Unification of Maxwell-Boltzmann, Bose-Einstein statistics and Zipf-Mandelbort Law with Law of Large Numbers and corresponding Fluctuation theorem

Vassili Kolokoltsov (v.kolokoltsov@warwick.ac.uk),
Tomasz M. Lapinski (84tomek@gmail.com)

University of Warwick,
Department of Statistics,
CV4 7AL Coventry, UK,

January 28, 2015

Abstract

The rigorous version of the result that unifies Maxwell-Boltzmann, Bose-Einstein statistics and Zipf-Mandelbrot Law is presented in the paper. Additionally, the Fluctuation theorem is included. Both results are set in the rigorous probabilistic framework. The system under consideration is composed of fixed number of particles and energy smaller than some prescribed value. The particles are redistributed on fixed number of energy levels and are indistinguishable within one level. Further, the degenerations of energy levels are included and their number increases as the number of particles increase. The three distribution mentioned in the title corresponds to the three asymptotic cases of rate of increases of degenerations as a function of particles. When we take a thermodynamic limit for three cases we basically obtain the Law of Large Numbers and the obtained three means are our distributions. The fluctuations, i.e. the deviation from the mean turns out depends on the Entropy of the system. When its maximum is in inside of the domain we get Gaussian distribution. When it is on the boundary we get discrete distribution in the direction orthogonal to the boundary on which maximum is situated and Gaussian in other directions. The proof is similar for both results. The main part are the properties and asymptotic behavior of Entropy. Due to this we can optimize it and therefore apply probabilistic results which underlying assumptions overlap with ours and eventually get the desired results.

Key words— Bose-Einstein, Maxwell-Boltzmann, Zipf-Mandelbrot, law of large numbers, fluctuations, Entropy

1 Introduction

Maxwell-Boltzmann and Bose-Einstein statistics are one of the most known results of Statistical Mechanics. The Zipf Law in turn, widely occurs in Complexity Science. For

this basic results we developed an interesting rigorous results which unifies them under one framework. Additionally we provide a corresponding fluctuation theorem and develop our results in the formal probabilistic manner.

The underlying probabilistic system is defined independent of the context of the application. However, for the illustrative purposes we consider physical context, namely, the ideal gas. We consider the system of number of particles which has an average energy of particle smaller or equal to some fixed value. It is similar to the microcanonical ensemble, but the energy is not fixed, it can have some prescribed range of values. However our results can be easily adopted to the case when value of energy is within some small bounded interval. Further, there are number of fixed energies which particles can have, i.e. energy levels and the number of these energy levels is also fixed. Each energy level has a number of level degenerations. These degenerations has an influence when counting possible number redistribution of particles over energy levels. The Entropy of the system is a product of logarithms of combinatoric formula counting the possible configuration on each energy level assuming indistinguishability of particles having the same energy.

There can be three types of system. We assume that the number of degenerations depends, increases, if the number of particles increase. We consider three cases depending on how strong this increase is. It can increase in the same way, slower or faster. This idea is introduced by Maslov [2005a]. For each of theses cases we obtain different statistics, respectively Bose-Einstein, Maxwell-Boltzmann and Zipf-Mandelbrot Law. Normally the degenerations correspond to the fact that there are energy levels which differs by very small energy values and therefore they are considered as the same level. In our case there is an open question what physical interpretation can have this changing number of degenerations.

The probability space which underlies the results is following. The sample space consists of the configuration of particles over the energy levels, divided by total number of particles. Theses configurations are constraint by fixed average energy and total number of particles. The sigma algebra is discrete and probability mass function of particular element in sample space is its entropy normalized by the partition function of the system.

Our first result is weak law of large numbers. The distribution of particles on energy levels converges to mean when the number of particles tends to infinity, i.e. in thermodynamic limit. The three distribution mentioned in the title are obtained for different rate of change of the degenerations number. Additionally we provide a rate of convergence to the mean. Similar result but not mathematically rigorous was introduced in Maslov [2005a] and Maslov [2005b]. The second result gives the distribution of the fluctuations from the mean as system tends to infinite size. Since the Entropy can have a maximum on the boundary or inside of the domain, depending on relation of fixed average energy of particle to average energy of energy levels, we have two cases. The idea of two cases was first mentioned in Maslov [2005a]. For the maximum of the Entropy in the interior of the domain we have Gaussain fluctuations. When the maximum is on the boundary we have discrete distribution in direction orthogonal to the boundary on which maximum is situated and Gaussian in other directions. Here we also provide a rate of convergence to the limiting distribution.

The proof for both results is similar. The essence of the proof are the properties of the Entropy and its asymptotic behaviour. Due to them we are able to optimize the Entropy and therefore distinguish the two types of its maximum and find the statistics explicitly. Similar calculations but partially rigorous and not complete were done in Maslov [2004]. The asymptotic properties of Entropy and the considered probabilistic system overlap

with the assumptions of the probabilistic results in Kolokoltsov and Lapinski [2015]. We apply the theorem from that paper for both types of maximum and for three cases of degenerations and obtain our limit theorems.

The paper is divided into several sections and appendix. The next section, second one is an introduction of the mathematical setting. We introduce there an underlying probabilistic system, a random variable and three cases of degenerations. In the third section we introduce and proof the main results of the paper, two limit theorems. Fourth section is devoted to the properties of Entropy, put as a Lemma with formal proof. The last section consists of the optimization of Entropy with related results needed in the optimizing. In the Appendix we provide some relatively basic results needed through out the paper.

2 Mathematical setting

For given integers $G, N > 0$, real number $E > 0$ and mapping $\varepsilon : \{1, 2, \ldots G\} \to \mathbb{R}$ we introduce a probability space. The elementary events are uniformly distributed G-dimensional vectors of nonnegative integers $n_i, i = 1, \ldots, G$ satisfying constraints:

$$N = n_1 + n_2 + \ldots + n_G, \tag{1}$$

$$EN \geq \varepsilon(1)n_1 + \varepsilon(2)n_2 + \ldots + \varepsilon(G)n_G. \tag{2}$$

In physics we call such system micro-canonical ensemble.

Arbitrary elementary event can be illustrated as the random distribution of N balls in G boxes. Moreover, each box has 'weight' coefficient $\varepsilon(i)$ and the total 'weight' must be less or equal EN.

Furthermore, let us denote the image of the function ε as the set $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m\}$ and without loss of generality it can be ordered $\varepsilon_1 < \varepsilon_2 < \ldots < \varepsilon_m$. To each element in the set corresponds a positive integer $G_i,\ i = 1, 2, \ldots, m$ representing the number of points in the domain of ε having the values ε_i, so that $G = \sum_{i=1}^m G_i$.

We can use this setting to define probability space in an alternative way. We consider the values G_i and $\varepsilon_i, i = 1, \ldots, m$ instead of the mapping ε. Respectively, the conditions (1) and (2) are reformulated

$$N = N_1 + N_2 + \ldots + N_m, \tag{3}$$

$$EN \geq \varepsilon_1 N_1 + \varepsilon_2 N_2 + \ldots + \varepsilon_m N_m, \tag{4}$$

where $N_i = n_{G_1+\ldots+G_{i-1}+1} + \ldots + n_{G_1+\ldots+G_{i-1}+2} + n_{G_1+\ldots+G_{i-1}+G_i}$ for $i = 1, \ldots, m$. This situation, can be illustrated as distributing N balls over m 'bigger' boxes, where to each corresponds unique value ε_i. Then in each i-th 'bigger' box balls are distributed over G_i boxes.

For given vectors $\mathcal{N} = (N_1, \ldots, N_m)$ and $\mathcal{G} = (G_1, \ldots, G_m)$ the number of different combinations which can occur in such redistribution, exactly the logarithm of that number is denoted by $S(\mathcal{N})$ and called Entropy.

We count those combinations using formula from Combinatorics for the possible number of unordered arrangements of size r obtained by drawing from n objects,

$$S(\mathcal{N}) = \ln \prod_{i=1}^{m} \frac{(N_i + G_i - 1)!}{N_i!(G_i - 1)!}. \tag{5}$$

Let us consider the discrete random vector denoted by $X_N = (X_1, X_2, \ldots, X_m)$ where $X_i = N_i/N,\ i = 1, \ldots, m$ and respectively sample space given by transformed conditions (3) and (4) is given by

$$\begin{aligned} &1 = x_1 + x_2 + \ldots + x_m, \\ &E \geq \varepsilon_1 x_1 + \varepsilon_2 x_2 + \ldots + \varepsilon_m x_m, \\ &x_i \in \left\{ \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1 \right\}, \end{aligned} \tag{6}$$

and denoted by $\Omega_{N,E}$ and respectively entropy function

$$S(x, N) = \ln \prod_{i=1}^{m} \frac{(x_i N + G_i - 1)!}{(x_i N)!(G_i - 1)!}.$$

The probability mass function (pmf) of random variable X_N is given by

$$Pr(X_N = x) = \frac{1}{Z(N, E)} e^{S(x,N)}, \tag{7}$$

where $Z(N, E)$ is a normalization constant specified by

$$Z(N, E) = \sum_{\Omega_{N,E}} e^{S(x,N)}, \tag{8}$$

which is a total number of elementary events in the sample space Ω_E. Sometimes $Z(N, E)$ is called partition function.

We are interested in the behaviour of random vector X as $N \to \infty$. Since the domain $\Omega_{N,E}$ depends on N the standard pointwise convergence of the function does not apply. However, we overcome that difficulty. For a fixed point x of the domain we choose a sequence of points $x(N)$ when $N \to \infty$. These points, for each N, are the closest points from the domain to the fixed point x. Since, the number of points of the domain increases with N and they are evenly distributed $x(N) \to x$. This way we obtain convergence for every fixed point of the domain and therefore have convergence analogical to pointwise convergence.

We consider a particular case when $G = G(N)$ is an increasing function of N. Moreover, for each N the components G_i are equally weighted and their number m remains constant. Which means that for all N, $G_i = g_i G(N)$ for $i = 1, \ldots, m$ and some constants g_i such that $\sum_{i=1}^{m} g_i = 1$.

We distinguish three cases of function $G(N)$, depending on its asymptotic behaviour in $N \to \infty$

$$\begin{aligned} &1)\ \frac{G(N)}{N} \to \infty, \\ &2)\ \frac{G(N)}{N} \to c, \\ &3)\ \frac{G(N)}{N} \to 0, \end{aligned} \tag{9}$$

where c is some positive constant. The idea of three asymptotic cases is adopted from the paper of Maslov Maslov [2005a].

3 Main results

Theorem 1 (Weak Law of large numbers). *Let X_N be the m-dimensional discrete random vector on the sample space $\Omega_{N,E}$ with pmf specified by (7). As $N \to \infty$ the random vector X_N converges in distribution to the constant vector $x^* = (x_1^*, x_2^*, \dots, x_m^*)$. The exact values of the components of x^* depend on the sample space parameter E.*

Let $\overline{g\varepsilon} = \frac{1}{m}\sum_{i=1}^m g_i\varepsilon_i$, then

a) When $\varepsilon_1 < E < \overline{g\varepsilon}$ the components of x^* are

$$
\begin{aligned}
&1)\ x_i^* = \frac{g_i}{e^{\lambda\varepsilon_i+\nu}}, && \text{if} \quad \frac{G(N)}{N} \to \infty,\\
&2)\ x_i^* = \frac{g_i}{e^{\lambda\varepsilon_i+\nu}-1}, && \text{if} \quad \frac{G(N)}{N} \to c,\\
&3)\ x_i^* = \frac{g_i}{\lambda\varepsilon_i+\nu}, && \text{if} \quad \frac{G(N)}{N} \to 0,
\end{aligned}
$$

for $i = 1, \dots, m$ and the parameters λ and ν are the solution of the system of equations

$$1 = \sum_{i=1}^m x_i^*, \;\; E = \sum_{i=1}^m \varepsilon_i x_i^*.$$

b) When $E \geq \overline{g\varepsilon}$ the components of x^* are

$$x_i^* = g_i, \;\; i = 1, \dots, m.$$

Further, we have following estimates for each case of $G(N)$ given by (9)

$$
\begin{aligned}
&1)\ M_{X_N}(\xi) = e^{\xi^T x^*} + O\left(\frac{1}{N^{1/2-3\epsilon}}\right), && \text{when } \frac{1}{N^{1/2-3\epsilon}} \gg \frac{N}{G(N)},\\
& M_{X_N}(\xi) = e^{\xi^T x^*} + O\left(\frac{N}{G(N)}\right), && \text{when } \frac{N}{G(N)} \gg \frac{1}{N^{1/2-3\epsilon}},
\end{aligned}
$$

$$
\begin{aligned}
&2)\ M_{X_N}(\xi) = e^{\xi^T x^*} + O\left(\frac{1}{N^{1/2-3\epsilon}}\right), && \text{when } \frac{1}{N^{1/2-3\epsilon}} \gg c(N),\\
& M_{X_N}(\xi) = e^{\xi^T x^*} + O(c(N)), && \text{when } c(N) \gg \frac{1}{N^{1/2-3\epsilon}},
\end{aligned}
$$

$$
\begin{aligned}
&3)\ M_{X_N}(\xi) = e^{\xi^T x^*} + O\left(\frac{1}{G(N)^{1/2-3\epsilon}}\right), && \text{when } \frac{1}{G(N)^{1/2-3\epsilon}} \gg \frac{G(N)}{N},\\
& M_{X_N}(\xi) = e^{\xi^T x^*} + O\left(\frac{G(N)}{N}\right), && \text{when } \frac{G(N)}{N} \gg \frac{1}{G(N)^{1/2-3\epsilon}},
\end{aligned}
$$

where $\epsilon \in (0, \min\{1/2m, 1/6\})$ is a constant, $N \to \infty$, $M_{X_N}(\xi)$ is a moment generating function of the random vector X_N. In the second case $\frac{G(N)}{N} = c + O(c(N))$, $c(N)$ is a positive, decreasing function and we define $f(x) \gg g(x) \iff \lim_{x\to\infty} f(x)/g(x) = \infty$.

Proof. Let us include the first constraint of domain $\Omega_{N,E}$ (6) by replacing the last component of x, $x_m = 1 - \sum_{i=1}^{m}$ in the function $S(x, N)$. The we can define a new $m-1$ random vector $X'_N = (X_1, \dots, X_{m-1})$, where its pmf is (7) with the substitution $x_m = 1 - \sum_{i=1}^{m} x_i$. The corresponding mgf of X'_N can be obtained from mgf of X_N

$$M_{X_N}(\xi) = E[e^{\xi^T X_N}] = E[e^{\xi'^T X'_N + \xi_m(1 - X_1 - \dots - X_{m-1})}] = E[e^{(\xi'^T - \xi_m)X'_N + \xi_m}] = e^{\xi_m} M_{X'_N}(\xi' - \xi_m) \tag{10}$$

Since the function $S(x, N)$ with $x_m = 1 - \sum_{i=1}^{m}$ has a properties given by the Lemma 1 of the Entropy results Section, therefore we can apply Theorem 1 from Kolokoltsov and Lapinski [2015]. We do it for all the cases of the $G(N)$ and obtain our theorem for the random vector X'_N except the exact vales of maximums. The estimates for the r.v. X'_N are following

$$1)\ M_{X'_N}(\xi' - \xi_m) = e^{(\xi' - \xi_m)^T x^*} + O\left(\frac{1}{N^{1/2 - 3\epsilon}}\right), \qquad \text{when } \frac{1}{N^{1/2-3\epsilon}} \gg \frac{N}{G(N)},$$
$$M_{X'_N}(\xi' - \xi_m) = e^{(\xi' - \xi_m)^T x^*} + O\left(\frac{N}{G(N)}\right), \qquad \text{when } \frac{N}{G(N)} \gg \frac{1}{N^{1/2-3\epsilon}},$$

$$2)\ M_{X'_N}(\xi' - \xi_m) = e^{(\xi' - \xi_m)^T x^*} + O\left(\frac{1}{N^{1/2 - 3\epsilon}}\right), \qquad \text{when } \frac{1}{N^{1/2-3\epsilon}} \gg c(N),$$
$$M_{X'_N}(\xi' - \xi_m) = e^{(\xi' - \xi_m)^T x^*} + O(c(N)), \qquad \text{when } c(N) \gg \frac{1}{N^{1/2-3\epsilon}},$$

$$3)\ M_{X'_N}(\xi' - \xi_m) = e^{(\xi' - \xi_m)^T x^*} + O\left(\frac{1}{G(N)^{1/2 - 3\epsilon}}\right), \qquad \text{when } \frac{1}{G(N)^{1/2-3\epsilon}} \gg \frac{G(N)}{N},$$
$$M_{X'_N}(\xi' - \xi_m) = e^{(\xi' - \xi_m)^T x^*} + O\left(\frac{G(N)}{N}\right), \qquad \text{when } \frac{G(N)}{N} \gg \frac{1}{G(N)^{1/2-3\epsilon}},$$

We obtain the estimate for our theorem by reversing the transformation (10) on the estimate

$$e^{\xi_m} e^{(\xi' + \xi_m)^T x^*} = e^{\xi'^T x^* - \xi_m(1 - x_1^* - \dots - x_{m-1}^*)} = e^{\xi'^T x^* - \xi_m x_{m-}^*} = e^{\xi^T x^*}$$

where x^* is m dimensional vector which is maximum of $S(x, N)$ in the limit as $N \to \infty$. The exact values of the maximum we obtain from the Lemma 2 of the Entropy results Section. Hence we obtained our theorem. □

For the discrete random vector X_N on the sample space $\Omega_{N,E}$ with pmf specified by (7) we define a m-dimensional random vector Y_N. We consider two cases of Y_N for two types of maximum of the function $S(x, N)$ which occurs in the pmf. Theses cases are defined by the Lemma 2 of the Entropy results section. The type of maximum is distinguished by the value of the sample space parameter E. The random vector Y_N has forms

$$a)\ Y_N = \sqrt{h(N)}(X_N - x^*), \qquad \text{when } E \geq \overline{ge}$$
$$b)\ Y_N = N(V_1 - v_1^*) + \sqrt{N}(\hat{V} - \hat{v}^*), \qquad \text{when } \epsilon_1 < E < \overline{ge} \tag{11}$$

where x^* and $x^* = Tv^* = T(v_1^*, \hat{v}^*)$ is the maximum of the functions S on Ω_N. The transformation $x = Tv$ is a rotation of the coordinate system such that the axis v_1 is orthogonal to the hyperplane of $\Omega_{N,E}$ on which maximum is attained. In the new corridanate system we have a transformed random vector $X_N = TV = T(V_1, \hat{V})$. Additionally, we assume that for the second type of maximum the hyperplane on which maximum is attained has a rational coefficient, i.e. values $\epsilon_1, \dots, \epsilon_m$ are rational.

Theorem 2 (Fluctuations theorem)**.** *As $N \to \infty$ the defined above random vector Y_N for the maximum type a) converges in distribution for all three cases of $G(N)$ to a $m-1$ dimensional random vector with normal distribution $\mathcal{N}(0, D^2 s(x^*)^{-1})$ and m-th random component Y_m is given by a combination of other random components $Y_m = -\sum_{i=1}^{m-1} Y_i$. For the maximum of type b) Y_N converges, for first two cases of $G(N)$, to a mixture of $m-2$-dimensional normal distribution $\mathcal{N}(0, \hat{D}^2 s(x^*)^{-1})$ along variable $\hat{V}$ and discrete distribution with pmf $\frac{\exp\{is'(x^*)\}}{\sum_{i=1}^{\infty}\exp\{is'(x^*)\}}$, where i is an index of points in the coordinate v_1 on the lattice $T(\Omega_N)$ starting form the maximal point v_1^* and m-th component of Y_N is given throught other components $Y_m = -\sum_{i=2}^{m-1} b_i Y_i$ in the limit, where b_i are some constants.*

The matrix of derivatives $D^2 s(x^*)^{-1}$ is $m-1 \times m-1$ and $\hat{D}^2 s(x^*)^{-1}$ is $m-2 \times m-2$ matrix along v_i, $i = 2, \dots, m$ and $s'(x^*)$ is derivative of s along v_1. Additionally we have the restriction, that the estimate is valid for a subsequence of integers N which elements can be divided by some integer q, where $\frac{p}{q} = v_1^*$ with $v^* = T^{-1}x^*$.
Furthermore, we have estimates for type a) of maximum

$$
\begin{aligned}
1)\ M_{Y_N}(\xi) &= e^{\frac{1}{2}\xi^T D^2 s(x^*)^{-1}\xi}\left(1 + O\left(\frac{1}{N^{1/2-3\epsilon}}\right)\right), && \text{when } \frac{1}{N^{1-3\epsilon}} \gg \frac{N}{G(N)},\\
M_{Y_N}(\xi) &= e^{\frac{1}{2}\xi^T D^2 s(x^*)^{-1}\xi}\left(1 + O\left(\frac{N^{3/2}}{G(N)}\right)\right), && \text{when } \frac{N}{G(N)} \gg \frac{1}{N^{1-3\epsilon}},
\end{aligned}
$$

$$
\begin{aligned}
2)\ M_{Y_N}(\xi) &= e^{\frac{1}{2}\xi^T D^2 s(x^*)^{-1}\xi}\left(1 + O\left(\frac{1}{N^{1/2-3\epsilon}}\right)\right), && \text{when } \frac{1}{N^{1-3\epsilon}} \gg c(N),\\
M_{Y_N}(\xi) &= e^{\frac{1}{2}\xi^T D^2 s(x^*)^{-1}\xi}\left(1 + O\big(\sqrt{G(N)}c(N)\big)\right), && \text{when } c(N) \gg \frac{1}{N^{1-3\epsilon}},
\end{aligned}
$$

$$
\begin{aligned}
3)\ M_{Y_N}(\xi) &= e^{\frac{1}{2}\xi^T D^2 s(x^*)^{-1}\xi}\left(1 + O\left(\frac{1}{G(N)^{1/2-3\epsilon}}\right)\right), && \text{when } \frac{1}{G(N)^{1-3\epsilon}} \gg \frac{G(N)}{N},\\
M_{Y_N}(\xi) &= e^{\frac{1}{2}\xi^T D^2 s(x^*)^{-1}\xi}\left(1 + O\left(\frac{G(N)^{3/2}}{N}\right)\right), && \text{when } \frac{G(N)}{N} \gg \frac{1}{G(N)^{1-3\epsilon}},
\end{aligned}
$$

and for the type b)

$$
\begin{aligned}
1)\ M_{Y_N}(\xi) &= \frac{\sum_{i=1}^{\infty} e^{is'(x^*)+\xi_1 i}}{\sum_{i=1}^{\infty} e^{is'(x^*)}} e^{\frac{1}{2}\hat{\xi}^T \hat{D}^2 s(x^*)^{-1}\hat{\xi}}\left(1 + O\left(\frac{1}{N^{1/2-3\epsilon}}\right)\right), && \text{when } \frac{1}{N^{1-3\epsilon}} \gg \frac{N}{G(N)},\\
M_{Y_N}(\xi) &= \frac{\sum_{i=1}^{\infty} e^{is'(x^*)+\xi_1 i}}{\sum_{i=1}^{\infty} e^{is'(x^*)}} e^{\frac{1}{2}\hat{\xi}^T \hat{D}^2 s(x^*)^{-1}\hat{\xi}}\left(1 + O\left(\frac{N^{3/2}}{G(N)}\right)\right), && \text{when } \frac{N}{G(N)} \gg \frac{1}{N^{1-3\epsilon}},
\end{aligned}
$$

$$
\begin{aligned}
2)\ M_{Y_N}(\xi) &= \frac{\sum_{i=1}^{\infty} e^{is'(x^*)+\xi_1 i}}{\sum_{i=1}^{\infty} e^{is'(x^*)}} e^{\frac{1}{2}\hat{\xi}^T \hat{D}^2 s(x^*)^{-1}\hat{\xi}}\left(1 + O\left(\frac{1}{N^{1/2-3\epsilon}}\right)\right), && \text{when } \frac{1}{N^{1-3\epsilon}} \gg c(N),\\
M_{Y_N}(\xi) &= \frac{\sum_{i=1}^{\infty} e^{is'(x^*)+\xi_1 i}}{\sum_{i=1}^{\infty} e^{is'(x^*)}} e^{\frac{1}{2}\hat{\xi}^T \hat{D}^2 s(x^*)^{-1}\hat{\xi}}\left(1 + O\big(\sqrt{G(N)}c(N)\big)\right), && \text{when } c(N) \gg \frac{1}{N^{1-3\epsilon}},
\end{aligned}
$$

as $N \to \infty$, where $\epsilon \in (0, \min\{1/2m, 1/6\})$ is some arbitrary small constant and M_{Y_N} is a moment generating function of random vector Y_N.

Proof. We start we the proof of the case a). We reduce the dimension of the underlying random vector of denoted by X_N. By using the constraint of the sample space (6) we reduce the dimension from m to $m-1$, i.e. apply the relation $x_m = 1-x_1-x_2-\ldots-x_{m-1}$. Then the first $m-1$ components of Y_m will be unchanged. The m-th component wil be equal

$$Y_m = \sqrt{N}(x_m - x_m^*) = \sqrt{N}(1 - \sum_{i=1}^{m-1} -x_m^*) = \sqrt{N}(\sum_{i=1}^{m} x_i^* - \sum_{i=1}^{m-1} x_i - x_m^*) = -\sum_{i=1}^{m-1} Y_i$$

where $x_m^* = 1 - \sum_{i=1}^{m-1}$.
Since the function $S(x, N)$ with $x_m = 1 - \sum_{i=1}^{m-1} x_i$ in the pmf (7) has properties given by the Lemma 1 in the Section on the Entropy results therefore the random vector $(Y_1, Y_2, \ldots, Y_{m-1})$ with considered pmf and the sample space $\Omega_{N,E}$ for the first two cases of $G(N)$ fulfils the requirements of the Theorem 2 from Kolokoltsov and Lapinski [2015]. We can apply that theorem and obtain results stated by this theorem.

For the case b) we again reduce dimension of the underlying random vector X_N. Since the Y_N can alternatively can be represented as

$$Y_N = NT_1^{-1}(x - x^*) + \sqrt{N}\hat{T}^{-1}(x - x^*)$$

where $T^{-1}x = v$ and the T_1^{-1} is 1-st row and $\hat{T}^{-1}$ is composed of rows $2, 3, \ldots, m$ of the matrix T^{-1}. If we set $x_m = 1 - \sum_{i=1}^{m-1} x_i$ then we can introduce altered transformation T' such that the random vector $(Y_1, Y_2, \ldots, Y_{m-1}) = Y_N'$ can be represented

$$Y_N' = N(v_1' - v'^*) + \sqrt{N}(\hat{v}' - \hat{v}'^*) = NT_1'^{-1}(x' - x'^*) + \sqrt{N}\hat{T}'^{-1}(x' - x'^*)$$

where $x' = (x_1, \ldots, x_{m-1})$ and prime generally denotes reduction in dimension. The we have for each Y_i

$$Y_1 = NT_1'^{-1}(x' - x'^*),$$
$$Y_j = \sqrt{N}T_j'^{-1}(x' - x'^*)$$

where $j = 2, \ldots, m-1$.
The component Y_m we can represent as

$$\begin{aligned}
Y_m &= \sqrt{N}T_m^{-1}(x - x^*) = \sqrt{N}\left(\sum_{i=1}^{m} t_{m,i}^{-1}(x_i - x_i^*)\right) = \sqrt{N}\left(\sum_{i=1}^{m-1} t_{m,i}^{-1}(x_i - x_i^*) + t_{m,m}^{-1}(x_m - x_m^*)\right) = \\
&= \sqrt{N}\left(\sum_{i=1}^{m-1} t_{m,i}^{-1}(x_i - x_i^*) + t_{m,m}^{-1}\left(1 - \sum_{i=1}^{m-1} x_i - x_m^*\right)\right) = \\
&= \sqrt{N}\left(\sum_{i=1}^{m-1} t_{m,i}^{-1}(x_i - x_i^*) - t_{m,m}^{-1}\sum_{i=1}^{m-1}(x_i - x_i^*)\right) = \sqrt{N}\sum_{i=1}^{m-1}(t_{m,i}^{-1} - t_{m,m}^{-1})(x_i - x_i^*) = \\
&= \sqrt{N}a^T(x' - x'^*)
\end{aligned}$$

where a, is some constant vector and $t_{j,i}^{-1}$ are components of T^{-1}. The resulting expression however can be represented as following linear combination

$$Y_m = \frac{b_1}{\sqrt{N}}Y_1 + \sum_{i=2}^{m-1} b_i Y_i,$$

where $b_1, b_2, \dots, b_m$ are some constants. In the limit , as $N \to \infty$ the first component tends to 0. Since the function $S(x, N)$ with $x_m = 1 - \sum_{i=1}^{m-1} x_i$ in the pmf (7) has properties given by the Lemma 1 in the Section on the Entropy Properties results therefore the random vector Y'_N with considered pmf and the sample space $\Omega_{N,E}$ for the three cases of $G(N)$ fulfils the requirements of the Theorem 2 from Kolokoltsov and Lapinski [2015]. We can apply that theorem and obtain results stated by this theorem.

□

4 Entropy properties

Lemma 1. *The Entropy $S(x, N)$ as $N \to \infty$ for each of the case of $G(N)$ given by (??), and for all $x \in \Omega_E$ has following properties*

$$\frac{\partial^2}{\partial x_i^2} S(x, N) < 0, \quad \frac{\partial^2}{\partial x_i \partial x_j} S(x, N) = 0, \text{ when } i \neq j, \text{ for all } N \tag{12}$$

$$\lim_{N\to\infty} \frac{\partial^2}{\partial x_i^2} S(x, N) < 0. \tag{13}$$

Additionally, for the first two cases of $G(N)$ and third respectively, we have

$$\begin{aligned} 1), 2)\ DS(x, N) =& N\big[Ds_l(x) + \sigma(x)\epsilon(N)\big],\ l = 1, 2, \\ 3)\ DS(x, N) =& G(N)\big[Ds_3(x) + \sigma(x)\epsilon(N)\big], \end{aligned} \tag{14}$$

as $N \to \infty$, where D is differential operator, $\sigma(x)$ some twice differentaible function of x and

$$1)\ s_1(x) = \sum_{i=1}^{m} \left[x_i \ln \frac{g_i}{x_i} + x_i \right], \tag{15}$$

$$2)\ s_2(x) = \sum_{i=1}^{m} \Big[(x_i + g_i c) \ln(x_i + g_i c) - x_i \ln x_i \Big], \tag{16}$$

$$3)\ s_3(x) = \sum_{i=1}^{m} \Big[g_i \ln x_i + g_i \Big]. \tag{17}$$

When $x_i = 0$ for some i-th component of x then corresponding summmad in the formula for $s_l, l = 1, 2, 3$ is equal to 0. The error term, $\epsilon(N)$ respectively for each $G(N)$ is defined

$$\begin{aligned}
1)\ \epsilon(N) &= O\left(\frac{1}{N}\right), & \text{when } \frac{1}{N} \gg \frac{N}{G(N)}, \\
\epsilon(N) &= O\left(\frac{N}{G(N)}\right), & \text{when } \frac{1}{N} \ll \frac{N}{G(N)}, \\
2)\ \epsilon(N) &= O\left(\frac{1}{N}\right), & \text{when } \frac{1}{N} \gg c(N), \\
\epsilon(N) &= O(c(N)), & \text{when } \frac{1}{N} \ll c(N), \\
3)\ \epsilon(N) &= O\left(\frac{1}{G(N)}\right), & \text{when } \frac{1}{G(N)} \gg \frac{G(N)}{N}, \\
\epsilon(N) &= O\left(\frac{G(N)}{N}\right), & \text{when } \frac{1}{G(N)} \ll \frac{G(N)}{N},
\end{aligned} \tag{18}$$

where in the second case $\frac{G(N)}{N} = c + O(c(N))$, $c(N)$ *is a positive, decreasing function and we define* $f(x) \gg g(x) \iff \lim_{x\to\infty} f(x)/g(x) = \infty$.
The properties of $S(x,N)$ *remain the same if we consider it as* $m-1$ *dimensional function with m-th component of* x *defined as* $x_m = 1 - \sum_{i=1}^{m-1} x_i$.

Proof. Since $\Gamma(N) = (N-1)!$ we can write

$$\frac{(x_iN + g_iG(N) - 1)!}{(x_iN)!(g_iG(N)-1)!} = \frac{\Gamma(x_iN + g_iG(N))}{\Gamma(x_iN+1)\Gamma(g_iG(N))},$$

for $i = 1, \ldots, m$. Further, let us introduce

$$\begin{aligned}
\Phi_i(N) &= -x_iN + g_iG(N) + \left(x_iN + g_iG(N) - \frac{1}{2}\right)\ln\big(x_iN + g_iG(N)\big),\\
\Psi_i(N) &= -x_iN - 1 + \left(x_iN + \frac{1}{2}\right)\ln\big(x_iN + 1\big),\\
\Theta_i(N) &= -g_iG(N) + \left(g_iG(N) - \frac{1}{2}\right)\ln g_iG(N).
\end{aligned} \tag{19}$$

First order approximation of gamma function by the Theorem 1 in the Appendix A, has asymptotic expansion

$$\Gamma(\lambda) \sim \sqrt{2\pi}e^{-\lambda+(\lambda+\frac{1}{2})\ln\lambda}\left[1 + \frac{1}{12\lambda} + \frac{1}{288\lambda^2} + \ldots\right], \lambda \to \infty.$$

For $x_i > 0$ expressions $x_iN+g_iG(N), x_iN+1, g_iG(N)$ are positive and increasing functions of N, the following Gamma functions can be approximated using the above asymptotic expansion

$$\begin{aligned}
\Gamma(x_iN + g_iG(N)) &\sim \sqrt{2\pi}e^{\Psi_i(N)}\left[1 + \frac{1}{12(x_iN + g_iG(N))} + \frac{1}{288(x_iN + g_iG(N))^2} + \ldots\right],\ x_iN + g_iG(N)\\
\Gamma(x_iN + 1) &\sim \sqrt{2\pi}e^{\Theta_i(N)}\left[1 + \frac{1}{12(x_iN+1)} + \frac{1}{288(x_iN+1)^2} + \ldots\right], x_iN + 1 \to \infty\\
\Gamma(g_iG(N)) &\sim \sqrt{2\pi}e^{\Phi_i(N)}\left[1 + \frac{1}{12g_iG(N)} + \frac{1}{288(g_iG(N))^2} + \ldots\right], g_iG(N) \to \infty,
\end{aligned}$$

where $\Phi_i(N), \Psi_i(N), \Theta_i(N)$ are given by (**??**).
From the definition of asymptotic expansion we obtain first order approximation, valid for $x_i > 0$ and corresponding large N

$$\begin{aligned}
\left|\Gamma(x_iN + g_iG(N)) - \sqrt{2\pi}e^{\Psi_i(N)}\right| &\le K_{i,\Psi}\left|\frac{1}{12(x_iN + g_iG(N))}\sqrt{2\pi}e^{\Psi_i(N)}\right|,\\
\left|\Gamma(x_iN + 1) - \sqrt{2\pi}e^{\Phi_i(N)}\right| &\le K_{i,\Phi}\left|\frac{1}{12(x_iN+1)}\sqrt{2\pi}e^{\Phi_i(N)}\right|,\\
\left|\Gamma(g_iG(N)) - \sqrt{2\pi}e^{\Theta_i(N)}\right| &\le K_{i,\Theta}\left|\frac{1}{12g_iG(N)}\sqrt{2\pi}e^{\Theta_i(N)}\right|,
\end{aligned}$$

where $K_{i,\Psi}, K_{i,\Phi}, K_{i,\Theta}$ are some positive constants. Note that we can increase these constants that above inequalities are also valid for $x_i = 0$ and all N. Hence we can represent

the above inequalities as

$$\Gamma(x_iN+g_iG(N))-\sqrt{2\pi}e^{\Psi_i(N)}=\frac{O(1)}{12(x_iN+g_iG(N))}\sqrt{2\pi}e^{\Psi_i(N)}, \tag{20}$$

$$\Gamma(x_iN+1)-\sqrt{2\pi}e^{\Phi_i(N)}=\frac{O(1)}{12(x_iN+1)}\sqrt{2\pi}e^{\Phi_i(N)}, \tag{21}$$

$$\Gamma(g_iG(N))-\sqrt{2\pi}e^{\Theta_i(N)}=\frac{O(1)}{12g_iG(N)}\sqrt{2\pi}e^{\Theta_i(N)}. \tag{22}$$

valid for $N\to\infty$. Next we combine approximations (22) and (21) using Lemma 6 from the Appendix A

$$\Gamma(x_iN+1)\Gamma(g_iG(N))-2\pi e^{\Psi_i(N)+\Theta_i(N)}=\frac{O(1)}{12(x_iN+1)12g_iG(N)}2\pi e^{\Psi_i(N)+\Theta_i(N)}+$$
$$+\frac{O(1)}{12(x_iN+1)}\sqrt{2\pi}e^{\Psi_i(N)+\Theta_i(N)}+\frac{O(1)}{12g_iG(N)}\sqrt{2\pi}e^{\Psi_i(N)+\Theta_i(N)},$$

which holds as $N\to\infty$.
Now depending on the asymptotic behaviour of $G(N)$ given by (**??**) this can be reexpressed as

$$1),2)\ \Gamma(x_iN+1)\Gamma(g_iG(N))-2\pi e^{\Psi_i(N)+\Theta_i(N)}=\frac{1}{N}\left[\frac{O(1)}{12(x_i+1/N)12g_iG(N)}2\pi e^{\Psi_i(N)+\Theta_i(N)}+\right.$$
$$\left.+\frac{O(1)}{12(x_i+1/N)}\sqrt{2\pi}e^{\Psi_i(N)+\Theta_i(N)}+\frac{O(1)}{12g_iG(N)/N}\sqrt{2\pi}e^{\Psi_i(N)+\Theta_i(N)}\right],$$
$$3)\ \Gamma(x_iN+1)\Gamma(g_iG(N))-2\pi e^{\Psi_i(N)+\Theta_i(N)}=\frac{1}{G(N)}\left[\frac{O(1)}{12(x_iN+1)12g_i}2\pi e^{\Psi_i(N)+\Theta_i(N)}+\right.$$
$$\left.+\frac{O(1)}{12(x_iN/G(N)+1/G(N))}\sqrt{2\pi}e^{\Psi_i(N)+\Theta_i(N)}+\frac{O(1)}{12g_i}\sqrt{2\pi}e^{\Psi_i(N)+\Theta_i(N)}\right],$$

and if we introduce constant $\sigma_{\Psi\Theta}(x)$ it can be simply stated as

$$1),2)\ \Gamma(x_iN+1)\Gamma(g_iG(N))-2\pi e^{\Psi_i(N)+\Theta_i(N)}=\frac{1}{N}\sigma_{\Psi\Theta}(x), \tag{23}$$
$$3)\ \Gamma(x_iN+1)\Gamma(g_iG(N))-2\pi e^{\Psi_i(N)+\Theta_i(N)}=\frac{1}{G(N)}\sigma_{\Psi\Theta}(x),$$

valid as $N\to\infty$, where

$$1),2)\ \sigma_{\Psi\Theta}(x)=\frac{O(1)}{12(x_i+1/N_0)12g_iG(N_0)}2\pi e^{\Psi_i(N_0)+\Theta_i(N_0)}+\frac{O(1)}{12(x_i+1/N_0)}\sqrt{2\pi}e^{\Psi_i(N_0)+\Theta_i(N_0)}+$$
$$+\frac{O(1)}{12g_iG(N_0)/N_0}\sqrt{2\pi}e^{\Psi_i(N_0)+\Theta_i(N_0)},$$
$$3)\ \sigma_{\Psi\Theta}(x)=\frac{O(1)}{12(x_iN_0+1)12g_i}2\pi e^{\Psi_i(N_0)+\Theta_i(N_0)}+\frac{O(1)}{12(x_iN_0/G(N)_0+1/G(N_0))}\sqrt{2\pi}e^{\Psi_i(N_0)+\Theta_i(N_0)}+$$
$$+\frac{O(1)}{12g_i}\sqrt{2\pi}e^{\Psi_i(N_0)+\Theta_i(N_0)},$$

Now use Lemma 7 from Appendix A to combine the asymptotic equations (20), (23) and after some basic manipulations obtain

$$\begin{aligned}
&1),2)\ \frac{\Gamma(x_iN+g_iG(N))}{\Gamma(x_iN+1)\Gamma(g_iG(N))}-\frac{\sqrt{2\pi}e^{\Phi_i(N)}}{2\pi e^{\Psi_i(N)+\Theta_i(N)}}\leq\\
&\leq\frac{\sqrt{2\pi}e^{\Phi_i(N)}}{2\pi e^{\Psi_i(N)+\Theta_i(N)}}\left(\frac{1}{1-\frac{1}{N}\sigma_{\Psi\Theta}(x)}\left(\frac{1}{N}\sigma_{\Psi\Theta}(x)+\frac{1}{12(x_iN+g_iG(N))}\right)\right),\\
&3)\ \frac{\Gamma(x_iN+g_iG(N))}{\Gamma(x_iN+1)\Gamma(g_iG(N))}-\frac{\sqrt{2\pi}e^{\Phi_i(N)}}{2\pi e^{\Psi_i(N)+\Theta_i(N)}}\leq\\
&\leq\frac{\sqrt{2\pi}e^{\Phi_i(N)}}{2\pi e^{\Psi_i(N)+\Theta_i(N)}}\left(\frac{1}{1-\frac{1}{G(N)}\sigma_{\Psi\Theta}(x)}\left(\frac{1}{G(N)}\sigma_{\Psi\Theta}(x)+\frac{1}{12(x_iN+g_iG(N))}\right)\right),
\end{aligned}$$

valid for $N>N_0$ and if we introduce constant $\sigma_{i,\Phi\Psi\Theta}(x)$ we can simply represent above inequalities as

$$\begin{aligned}
&1),2)\ \frac{\Gamma(x_iN+g_iG(N))}{\Gamma(x_iN+1)\Gamma(g_iG(N))}-\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)}=\frac{1}{N}\frac{\sigma_{i,\Phi\Psi\Theta}(x)}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)},\\
&3)\ \frac{\Gamma(x_iN+g_iG(N))}{\Gamma(x_iN+1)\Gamma(g_iG(N))}-\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)}=\frac{1}{G(N)}\frac{\sigma_{i,\Phi\Psi\Theta}(x)}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)},
\end{aligned}\tag{24}$$

valid as $N\to\infty$ with

$$\begin{aligned}
&1),2)\ \sigma_{i,\Phi\Psi\Theta}(x)=\frac{1}{1-\frac{1}{N_0}\sigma_{\Psi\Theta}(x)}\left(\sigma_{\Psi\Theta}(x)+\frac{1}{12(x_i+g_iG(N_0)/N_0)}\right),\\
&3)\ \sigma_{i,\Phi\Psi\Theta}(x)=\frac{1}{1-\frac{1}{G(N_0)}\sigma_{\Psi\Theta}(x)}\left(\frac{1}{G(N)}\sigma_{\Psi\Theta}(x)+\frac{1}{12(x_iN_0/G(N_0)+g_i)}\right),
\end{aligned}$$

when N_0 is some large constant. Now we use Lemma 6 from the Appendix A to get expression for the m factors for the first and second case of $G(N)$

$$\begin{aligned}
&\prod_{i=1}^{m}\frac{\Gamma(x_iN+g_iG(N))}{\Gamma(x_iN+1)\Gamma(g_iG(N))}-\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)}=\\
&=\prod_{i=1}^{m}\frac{1}{N}\sigma_{i,\Phi\Psi\Theta}(x)\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)}+\sum_{j=1}^{m-1}\sum_{C_m^{m-j,j}}\prod_{k=1}^{m-j}\prod_{l=m-j+1}^{m}\frac{1}{N}\sigma_{k,\Phi\Psi\Theta}(x)\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)},
\end{aligned}$$

valid as $N\to\infty$ and if we introduce constant $\sigma_{\Phi\Psi\Theta}(x)$ we can simplify above asymptotic inequality for all cases of $G(N)$ to

$$\begin{aligned}
&1),2)\ \prod_{i=1}^{m}\frac{\Gamma(x_iN+g_iG(N))}{\Gamma(x_iN+1)\Gamma(g_iG(N))}-\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)}=\frac{1}{N}\sigma_{\Phi\Psi\Theta}(x)\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)},\\
&3)\ \prod_{i=1}^{m}\frac{\Gamma(x_iN+g_iG(N))}{\Gamma(x_iN+1)\Gamma(g_iG(N))}-\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(N)}=\frac{1}{G(N)}\sigma_{\Phi\Psi\Theta}(x)\prod_{i=1}^{m}\frac{1}{\sqrt{2\pi}}e^{\Phi_i(N)-\Psi_i(N)-\Theta_i(
\end{aligned}\tag{25}$$

where

$$1),2)\ \sigma_{\Phi\Psi\Theta}(x) = \frac{1}{N_0^{m-1}}\prod_{i=1}^{m}\sigma_{i,\Phi\Psi\Theta}(x) + \sum_{j=1}^{m-1}\sum_{C_m^{m-j,j}}\prod_{k=1}^{m-j}\frac{1}{N_0^{k-1}}\prod_{l=m-j+1}^{m}\sigma_{i_k,\Phi\Psi\Theta}(x),$$

$$3)\ \sigma_{\Phi\Psi\Theta}(x) = \frac{1}{G(N_0)^{m-1}}\prod_{i=1}^{m}\sigma_{i,\Phi\Psi\Theta}(x) + \sum_{j=1}^{m-1}\sum_{C_m^{m-j,j}}\prod_{k=1}^{m-j}\frac{1}{G(N_0)^{k-1}}\prod_{l=m-j+1}^{m}\sigma_{i_k,\Phi\Psi\Theta}(x).$$

Now, let us rearrange asymptotic equation (25) in terms of the function $S(x,N)$

$$1),2)\ \exp[S(x,N)] = \exp\left[\ln(2\pi)^{-m/2} + \sum_{i=1}^{m}\Big(\Phi_i(N) - \Psi_i(N) - \Theta_i(N)\Big) + \ln\left(1 + \frac{1}{N}\sigma_{\Phi\Psi\Theta}(x)\right)\right],$$

$$3)\ \exp[S(x,N)] = \exp\left[\ln(2\pi)^{-m/2} + \sum_{i=1}^{m}\Big(\Phi_i(N) - \Psi_i(N) - \Theta_i(N)\Big) + \ln\left(1 + \frac{1}{G(N)}\sigma_{\Phi\Psi\Theta}(x)\right)\right].$$

Since $\ln(1+x) = O(x)$ when $x \to 0$ we can simplify above equations into

$$1),2)\ S(x,N) = \ln(2\pi)^{-m/2} + \sum_{i=1}^{m}\Big(\Phi_i(N) - \Psi_i(N) - \Theta_i(N)\Big) + \frac{1}{N}\sigma_S(x), \tag{26}$$

$$3)\ S(x,N) = \ln(2\pi)^{-m/2} + \sum_{i=1}^{m}\Big(\Phi_i(N) - \Psi_i(N) - \Theta_i(N)\Big) + \frac{1}{G(N)}\sigma_S(x),$$

as $N \to \infty$ where $\sigma_S(x) = O(\sigma_{\Phi\Psi\Theta}(x))$.

Now, let us prove first result of the lemma. We start by explicitly evaluating functions Φ, Ψ and Θ using (19) for the first case of $G(N)$

$$\begin{aligned} S(x,N) =& \ln(2\pi)^{-m/2} + \sum_{i=1}^{m}\left[\left(x_i(N) + g_iG(N) - \frac{1}{2}\right)\ln(x_iN + g_iG(N)) - \right. \\ & \left.\left(x_iN + \frac{1}{2}\right)\ln(x_iN+1) - \left(g_iG(N) - \frac{1}{2}\right)\ln g_iG(N) + 1\right] + \frac{1}{N}\sigma_S(x). \end{aligned} \tag{27}$$

valid for $x_i \in [0,1]$ and large enough N.
Then we calculate second derivative along x_i, for some i within avaiable range

$$\frac{\partial^2}{\partial x_i^2}S(x,N) = \left[\frac{2N^2}{x_iN + g_iG(N)} - (x_iN + g_iG(N) - 1/2)\frac{N^2}{(x_iN + g_iG(N))^2} - \right. \tag{28}$$

$$\left. - \frac{2N^2}{x_iN+1} + (x_iN + 1/2)\frac{N^2}{(x_iN+1)^2}\right] + \frac{1}{N}\frac{\partial^2\sigma_S(x_i)}{\partial x_i^2}. \tag{29}$$

For $x_i = 0$ the second derivative becomes

$$\frac{\partial^2}{\partial x_i^2}S(x,N) = \ln(2\pi)^{-m/2}\left[\frac{N^2}{g_iG(N)} - \frac{1}{2}\frac{N^2}{(g_iG(N))^2} - \frac{3}{2}N^2\right] + \frac{1}{N}\frac{\partial^2\sigma_S(0)}{\partial x_i^2}.$$

which is negative for any case of $G(N)$ and large N. Note, that σ_S when $x_i = 0$ becomes a constant. We prove for $x_i \neq 0$ be transforming the second derivative (28)

$$\begin{aligned} \frac{\partial^2}{\partial x_i^2}S(x,N) =& \ln(2\pi)^{-m/2}\left[-\frac{N^2(g_iG(N)-1)}{(x_iN + g_iG(N))(x_iN+1)} - \frac{N^2}{(x_iN+1)^2} + \right. \\ & \left. + \frac{N^2}{(x_iN + g_iG(N))^2}\right] + \frac{1}{N}\frac{\partial^2\sigma_S(x_i)}{\partial x_i^2} < 0, \end{aligned}$$

for large enough N. Note the for the third case of $G(N)$ we get the same result. The difference lies only in the change from N to $G(N)$ in the last expression which does not have influence for large N.
Further, it is immidiet to see that in the limit as $N \to \infty$, when $x_i = 0$ the sign of second derivative is unchanged. For positive values of x_i we transform the second derivative to the form for three cases of $G(N)$

$$\begin{aligned}
1), 2)\ \frac{\partial^2}{\partial x_i^2} S(x, N) =& N\Bigg[-\frac{(g_i - 1/G(N))}{(x_i N/G(N) + g_i)(x_i + 1/N)} - \frac{1}{N(x_i + 1/N)^2} + \\
&+ \frac{1}{N(x_i + g_i G(N)/N)^2}\Bigg] + \frac{1}{N}\frac{\partial^2 \sigma_S(x_i)}{\partial x_i^2} \to -\infty, \\
3)\ \frac{\partial^2}{\partial x_i^2} S(x, N) =& G(N)\Bigg[-\frac{(g_i - 1/G(N))}{(x_i + g_i G(N)/N)(x_i + 1/N)} - \frac{1}{G(N)(x_i + 1/N)^2} + \\
&+ \frac{1}{G(N)(x_i + g_i G(N)/N)^2}\Bigg] + \frac{1}{G(N)}\frac{\partial^2 \sigma_S(x_i)}{\partial x_i^2} \to -\infty,
\end{aligned}$$

Hence we get the first results of the lemma.

To get the last results of the lemma we first further transform the approximation of Entropy

$$\begin{aligned}
S(x, N) =& \ln(2\pi)^{-m/2} + \sum_{i=1}^{m}\Bigg[\big(x_i(N) + g_i G(N)\big)\ln(x_i N + g_i G(N)) - x_i N \ln x_i N - \\
&- g_i G(N)\ln g_i G(N) - \frac{1}{2}\ln(x_i N + g_i G(N)) - x_i N \ln\left(1 + \frac{1}{x_i N}\right) - \\
&- \frac{1}{2}\ln(x_i N + 1) + \frac{1}{2}\ln g_i G(N) + 1\Bigg] + \frac{1}{N}\sigma_S(x).
\end{aligned} \tag{30}$$

for the first case of $G(N)$.
Now, for each case of $G(N)$ we perform calculations separately.

1) We further transform the Entropy and eventually obtain

$$\begin{aligned}
S(x, N) =& \ln(2\pi)^{-m/2} + N\sum_{i=1}^{m}\Bigg[x_i \ln\frac{g_i}{x_i} + \left(x_i + g_i\frac{G(N)}{N}\right)\ln\left(1 + \frac{x_i N}{g_i G(N)}\right) - \\
&- \frac{1}{2N}\ln(x_i N + g_i G(N)) - x_i \ln\left(1 + \frac{1}{x_i N}\right) - \frac{1}{2N}\ln(x_i N + 1) + \frac{1}{2N}\ln g_i G(N) + \frac{1}{N}\Bigg] + \\
&+ N\ln\frac{G(N)}{N} + \frac{1}{N}\sigma_S(x),
\end{aligned}$$

where during transformation we substituted the constraint of the domain Ω_E (**??**). Then we approximate appropriate logarithm using $\ln(1 + x) = x + O(x^2), x \to 0$ to get isolate the function $s_1(x)$

$$\begin{aligned}
S(x, N) =& \ln(2\pi)^{-m/2} + N\sum_{i=1}^{m}\Bigg[x_i \ln\frac{g_i}{x_i} + x_i + \frac{O(1)x_i N}{g_i^2 G(N)} + x_i \ln\left(1 + \frac{x_i N}{g_i G(N)}\right) - \\
&- \frac{1}{2N}\ln(x_i N + g_i G(N)) - x_i \ln\left(1 + \frac{1}{x_i N}\right) - \frac{1}{2N}\ln(x_i N + 1) + \frac{1}{2N}\ln g_i G(N) + \frac{1}{N}\Bigg] + \\
&+ N\ln\frac{G(N)}{N} + \frac{1}{N}\sigma_S(x),
\end{aligned}$$

The first two terms inside the bracket are components of the function $s(x)$ other tend to 0. Expressions outside the bracket depends on N. Then we calculate the derivative along x_i of the previously obtained Entropy and get

$$\begin{aligned}S'(x,N) =& N\left[\ln\frac{g_i}{x_i}+\ln\left(1+\frac{x_iN}{g_iG(N)}\right)-\frac{1}{2}\frac{1}{x_iN+g_iG(N)}-\right.\\&\left.-\ln\left(1+\frac{1}{x_iN}\right)+\frac{1}{2x_iN+2}\right]+\frac{1}{N}\sigma'_S(x).\end{aligned}$$

Then, we perform some convenient rearrangements of expressions

$$\begin{aligned}S'(x,N) =& Ns_1'(x)+N\frac{N}{G(N)}\left[\frac{G(N)}{N}\ln\left(1+\frac{x_iN}{g_iG(N)}\right)\right]+\\&+N\frac{1}{N}\left[\sum_{i=1}^{m-1}\left(-\frac{1}{2}\frac{1}{x_i+g_iG(N)/N}-N\ln\left(1+\frac{1}{x_iN}\right)+\frac{1}{2x_i+2/N}\right)\right],\end{aligned}$$

with

$$s_1'(x)=\ln\frac{g_i}{x_i}.$$

where the logarithms in the expression are bounded by a constant, independed of N. This can be seen when we apply the expansion of logarithm $\log(1+x)=x+O(x^2)$ when $x\to 0$ to those logarithms.

We can write above formula in the following way

$$\begin{aligned}S'(x,N)-Ns_1'(x) &= K_{1,i}(x,N)\frac{1}{N}, &&\text{when } \frac{1}{N}\gg\frac{N}{G(N)},\\ S'(x,N)-Ns_1'(x) &= K_{1,i}(x,N)\frac{N}{G(N)}, &&\text{when } \frac{1}{N}\ll\frac{N}{G(N)},\end{aligned}$$

or alternatively with constants $\sigma(x)$

$$\begin{aligned}S'(x,N) &= N\left[s_1'(x)+\sigma(x)\frac{1}{N}\right], &&\text{when } \frac{1}{N}\gg\frac{N}{G(N)},\\ S'(x,N) &= N\left[s_1'(x)+\sigma(x)\frac{N}{G(N)}\right], &&\text{when } \frac{1}{N}\ll\frac{N}{G(N)},\end{aligned}$$

valid for $N>N_0$, where $\sigma(x)$ are defined for two cases

$$\sigma(x) =O\left(\frac{N_0^2}{G(N_0)}\left[\frac{G(N_0)}{N_0}\ln\left(1+\frac{x_iN_0}{g_iG(N_0)}\right)\right]+\left(\frac{1}{2}\frac{1}{x_i+g_iG(N_0)/N_0}+\frac{3}{2x_i}+O\left(\frac{1}{x_i^2N_0}\right)\right)\right),$$

$$\sigma(x) =O\left(\frac{G(N_0)}{N_0}\ln\left(1+\frac{x_iN_0}{g_iG(N_0)}\right)\right]+\frac{G(N_0)}{N_0^2}\left[\sum_{i=1}^{m-1}\left(\frac{1}{2}\frac{1}{x_i+g_iG(N_0)/N_0}+\frac{3}{2x_i}+O\left(\frac{1}{x_i^2N_0}\right)\right)\right]\right).$$

This result can be asly extended to the vector of derivatives $DS(x,N)$.
Hence we get the first case of the second result of the lemma.

2) For this case we also further transform (??)

$$S(x,N) = \ln(2\pi)^{-m/2} + N\sum_{i=1}^{m}\left[\left(x_i + g_i\frac{G(N)}{N}\right)\ln\left(x_i + g_i\frac{G(N)}{N}\right) - x_i\ln x_i - g_i\frac{G(N)}{N}\ln g_i\frac{G(N)}{N} - \right.$$
$$\left. - \frac{1}{2N}\ln(x_iN + g_iG(N)) - x_i\ln\left(1+\frac{1}{x_iN}\right) - \frac{1}{2N}\ln(x_iN+1) + \frac{1}{2N}\ln g_iG(N) + \frac{1}{N}\right] +$$
$$+ \frac{1}{N}\sigma_S(x).$$

In order to isolate the function $s_2(x)$ we use asymptotic inequality $\frac{G(N)}{N} = c + O(c(N))$, where $c(N)$ is a positive, decreasing function

$$S(x,N) = \ln(2\pi)^{-m/2} + N\sum_{i=1}^{m}\left[\left(x_i + g_ic\right)\ln\left(x_i+g_ic\right) + \left(x_i+g_ic\right)\ln\left(1+\frac{g_ic(N)}{x_i+g_ic}\right) + \right.$$
$$+ g_ic(N)\ln\left(x_i + g_i\frac{G(N)}{N}\right) - x_i\ln x_i - g_i\frac{G(N)}{N}\ln g_i\frac{G(N)}{N} -$$
$$\left. - \frac{1}{2N}\ln(x_iN + g_iG(N)) - x_i\ln\left(1+\frac{1}{x_iN}\right) - \frac{1}{2N}\ln(x_iN+1) + \frac{1}{2N}\ln g_iG(N) + \frac{1}{N}\right] +$$
$$+ \frac{1}{N}\sigma_S(x).$$

Then we calculate the derivative along x_i of the previously obtained Entropy and get

$$S'(x,N) = \ln(2\pi)^{-m/2} + N\sum_{i=1}^{m}\left[\ln\left(1+\frac{g_iG(N)}{x_iN}\right) + \ln\left(1+\frac{x_iN}{g_iG(N)}\right) - \frac{1}{2}\frac{1}{x_iN+g_iG(N)} - \right.$$
$$\left. - \ln\left(1+\frac{1}{x_iN}\right) + \frac{1}{2x_iN+2}\right] + \frac{1}{N}\sigma'_S(x).$$

Then, as for the first case, we perform some convenient rearrangements of expressions

$$S'(x,N) = \ln(2\pi)^{-m/2} + N\sum_{i=1}^{m}\left[c(N)\frac{O(1)g_i}{x_i(1+\frac{g_i}{x_i}c + x_{\theta(N)})} + \frac{1}{N}\left(-\frac{1}{2}\frac{1}{x_i + g_iG(N)/N} - \right.\right.$$
$$\left.\left. - N\ln\left(1+\frac{1}{x_iN}\right) + \frac{1}{2x_i + 2/N}\right)\right] + \frac{1}{N}\sigma'_S(x).$$

where the logarithms in the expression are bounded by a constant, independent of N. This can be seen when we apply the expansion of logarithm. Note that we have used $\frac{G(N)}{N} = c + O(c(N))$ in the transformation. We can write above formula in the following way

$$S'(x,N) - Ns'_2(x) = K_{2,i}(x,N)\frac{1}{N}, \qquad \text{when } \frac{1}{N} \gg c(N),$$
$$S'(x,N) - Ns'_2(x) = K_{2,i}(x,N)c(N), \qquad \text{when } \frac{1}{N} \ll c(N),$$

or alternatively with constants $\sigma(x)$

$$S'(x,N) = N\left[s'_2(x) + \sigma(x)\frac{1}{N}\right], \qquad \text{when } \frac{1}{N} \gg c(N),$$
$$S'(x,N) = N\left[s'_2(x) + \sigma(x)c(N)\right], \qquad \text{when } \frac{1}{N} \ll c(N),$$

valid for $N > N_0$, with

$$\sigma(x) = O\left(N_0 c(N_0) \frac{O(1) g_i}{x_i(1+\frac{g_i}{x_i}c)} + \frac{1}{2}\frac{1}{x_i + g_i G(N_0)/N_0} + \frac{3}{2x_i} + O\left(\frac{1}{x_i^2 N_0}\right)\right),$$

$$\sigma(x) = O\left(\frac{O(1) g_i}{x_i(1+\frac{g_i}{x_i}c)} + \frac{1}{c(N_0)N_0}\left(\frac{1}{2}\frac{1}{x_i + g_i G(N_0)/N_0} + \frac{3}{2x_i} + O\left(\frac{1}{x_i^2 N_0}\right)\right)\right),$$

as $N \to \infty$ and the second case is proved.

3) First we transform (**??**) for the third case, i.e. the error has $1/G(N)$ instead of $1/N$

$$S(x,N) = \ln(2\pi)^{-m/2} + G(N)\sum_{i=1}^{m}\left[g_i \ln x_i + \left(x_i \frac{N}{G(N)} + g_i\right)\ln\left(1+\frac{g_i G(N)}{x_i N}\right) - g_i \ln g_i G(N) + \right.$$
$$+ g_i \ln N - \frac{1}{2G(N)}\ln(x_i N + g_i G(N)) - x_i\frac{N}{G(N)}\ln\left(1+\frac{1}{x_i N}\right) - \frac{1}{2G(N)}\ln(x_i N + 1) +$$
$$\left.\frac{1}{2G(N)}\ln g_i G(N) + \frac{1}{G(N)}\right] + \frac{1}{G(N)}\sigma_S(x).$$

The function $s_3(x)$ can be isolated by applying logarithm expansion to the second term in the bracket. Then we differentiate along x_i and perform appropriate rearrangements

$$S'(x,N) = \ln(2\pi)^{-m/2} + G(N)\sum_{i=1}^{m}\left[\frac{G(N)}{N}O\left(\frac{g_i^2 G(N)}{x_i^2 N}\right) + \frac{1}{G(N)}\left(-\frac{1}{2}\frac{1}{x_i + g_i G(N)/N} - \right.\right.$$
$$\left.\left. - \frac{1}{x_i} - O\left(\frac{1}{x_i^2 N}\right) + \frac{1}{2x_i + 2/N}\right)\right] + \frac{1}{G(N)}\sigma'_S(x).$$

which can be represented as

$$S'(x,N) = N\left[s'_3(x) + \sigma(x)\frac{1}{G(N)}\right], \qquad \text{when } \frac{1}{G(N)} \gg \frac{G(N)}{N},$$
$$S'(x,N) = N\left[s'_3(x) + \sigma(x)\frac{G(N)}{N}\right], \qquad \text{when } \frac{1}{G(N)} \ll \frac{G(N)}{N},$$

valid for $N > N_0$, where $\sigma(x)$ are defined

$$\sigma(x) = O\left(\frac{N_0^2}{G(N_0)}O\left(\frac{g_i^2 G(N_0)}{x_i^2 N_0}\right) + \frac{1}{2}\frac{1}{x_i + g_i G(N_0)/N_0} + \frac{3}{2x_i} + O\left(\frac{1}{x_i^2 N_0}\right)\right),$$

$$\sigma(x) = O\left(O\left(\frac{g_i^2 G(N_0)}{x_i^2 N_0}\right) + \frac{N_0}{G(N_0)^2}\left(\frac{1}{2}\frac{1}{x_i + g_i G(N_0)/N_0} + \frac{3}{2x_i} + O\left(\frac{1}{x_i^2 N_0}\right)\right)\right).$$

Hence we get the error terms $\epsilon(N)$ and last result is proved.

□

5 Entropy optimization

We consider maximization problem of the function $S(x,N)$ over the domain Ω_E for all N and in the limit, as $N \to \infty$.

Lemma 2. *The function $S(x,N)$ for every N, when parameter of the domain $E > \varepsilon_1$, has a unique maximal point denoted by $x^*(N)$. As $N \to \infty$ the maximum exists only if the parameter of the domain, $E > \varepsilon_1$ and it is unique.*
For $\overline{g\varepsilon} = \frac{1}{m}\sum_{i=1}^m g_i\varepsilon_i$, if $E \geq \overline{g\varepsilon}$, then the maximal point $x^* = (x_1^*, x_2^*, \ldots, x_m^*)$ has components

$$x_i^* = g_i, i = 1, \ldots, m,$$

and if $\varepsilon_1 < E < \overline{g\varepsilon}$, respectively for each case of $G(N)$, the maximal point has components

$$\begin{aligned} &1)\ x_i^* = \frac{g_i}{e^{\lambda\varepsilon_i+\nu}}, \\ &2)\ x_i^* = \frac{g_i c}{e^{\lambda\varepsilon_i+\nu}-1}, \\ &3)\ x_i^* = \frac{g_i}{\lambda\varepsilon_i+\nu}, \end{aligned}$$

for $i = 1, \ldots, m$, where the parameters λ, ν are uniquely determined by the equations

$$\sum_{i=1}^m x_i^* = 1, \quad \sum_{i=1}^m \varepsilon_i x_i^* = E.$$

Proof. The domain Ω_E is an intersection of two m-dimensional simplexes, first is determined by origin and standard basis vectors of $\mathbb{R}^m$, i.e., $o, e_1, \ldots, e_m$ and second by vectors $0, \varepsilon_1 e_1/E, \ldots, \varepsilon_m e_m/E$. Since simplexes are convex sets, so their intersection and therefore the domain is convex. Since the function $S(x,N)$ for all $x \in \Omega_E$ and N has properties (12), hence its matrix of second order derivatives $D^2S(x,N)$ is diagonal matrix with negative entries. Therefore it is a strictly concave function on a convex domain. When the domain is nonempty, $E > \varepsilon_1$, the unique maximum exists and we denote it by $x^*(N)$.

In the limit $N \to \infty$, in order to get maximum x^* independent of N we consider function $\lim_{N\to\infty} \frac{1}{N}S(x,N)$. Then the properties (13) and (14) of S implies

$$\lim_{N\to\infty} \frac{1}{N}\frac{\partial^2 S(x,N)}{\partial x_i^2} < 0$$

for all $x \in \Omega_E$. From (12) we infer that in the limit the mixed derivatives are equal to 0. Hence the matrix of second derivatives $\lim_{N\to\infty} \frac{1}{N}D^2S(x,N)$ is a diagonal matrix with negative entries. Therefore is negative definite matrix and $\frac{1}{N}S(x,N)$ is a strictly concave function in the limit.

Since considered function is strictly concave, defined on a convex domain Ω_E, constraints (**??**) and (**??**) are affine, it is a convex optimization problem, for definition and terminology see Appendix A.4. Since objective is strictly concave the optimal vector, denoted as x^* is unique, if exists.

Now, we find the explicit form of the optimal vector. We start by representing our optimization problem in so called 'standard form'

$$\begin{aligned} \text{minimize} \quad & -f_l, \\ \text{subject to} \quad & \varepsilon^T x - E \leq 0, \\ & \mathbf{1}^T x - 1 = 0, \end{aligned} \tag{31}$$

where $\mathbf{1}$ is the unit vector and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)$.

Since the considered problem is convex and if $E > \varepsilon_1$ there exists $x \in \text{Relint}(\Omega)$, (for the definition of Relint see Appendix A.4) with

$$\begin{aligned} \mathbf{1}^T x - 1 &= 0, \\ \varepsilon^T x - E &< 0, \end{aligned}$$

the Slater's conditions holds (see Appendix on Theory of Optimization A.4) therefore strong duality occurs and optimal point exists. Then for each of the three cases of $G(N)$ we have corresponding Karush-Kuhn-Tucker (KKT) conditions (for details see Appendix A.4),

$$\begin{aligned} \varepsilon^T x^* - E &\leq 0, \\ \mathbf{1}^T x^* - 1 &= 0, \\ \lambda &\geq 0, \\ \lambda\Big(\varepsilon^T x^* - E\Big) &= 0, \\ -\lim_{N\to\infty} \frac{1}{N} DS(x^*, N) + \lambda^* \varepsilon + \nu \mathbf{1} &= o, \end{aligned} \tag{32}$$

For the convex optimization problem with the strong duality, these are necessary and sufficient conditions for the vectors x^* and (λ, ν) to be optimal.

Now, we solve (32). Since

$$\lim_{N\to\infty} \frac{1}{N} DS(x, N) = s_l(x)$$

with $l = 1, 2, 3$ respectively for each case of $G(N)$. Hence for the first function $s_1(x) = \sum_{i=1}^m x_i \ln \frac{g_i}{x_i} + x_i$ from the last of KKT conditions we obtain

$$x_i^* = \frac{g_i}{e^{\lambda E_i + \nu}} \tag{33}$$

Next, first four equations we represent as two cases for possible ranges of values of λ. The first case is

$$\begin{aligned} \varepsilon^T x^* - E &\leq 0, \\ \mathbf{1}^T x^* - 1 &= 0, \\ \lambda &= 0. \end{aligned}$$

From Lemma 3 attached at the end of this section, the solution for the above system exists and is unique only if $E \geq \overline{g\varepsilon}$ and $(\lambda, \nu) = (0, 0)$. Hence (33) becomes

$$x_i = g_i, i = 1, \ldots, m.$$

The other case is

$$\begin{aligned} \varepsilon^T x^* - E &= 0, \\ \mathbf{1}^T x^* - 1 &= 0, \\ \lambda &> 0, \end{aligned} \tag{34}$$

where we have equality in the first condition because of $\lambda^*\Big(\varepsilon^T x^* - E\Big) = 0$. By Lemma 3 of this section, solution (λ, ν) exists and is unique only if $\varepsilon_1 < E < \overline{g\varepsilon}$. Further, substituting (33) into two first conditions of (34) and we get the system of equations from which we can calculate parameters λ and ν explicitly. For the second and third case, situation is analogical but we use respectively Lemma 5 and 4. Only the outcome of the last KKT condition gives different result. □

5.1 Related results

For the positive numbers $g_i, \varepsilon_i, i = 1, \ldots, m$ such that $\varepsilon_1 < \varepsilon_2 < \ldots < \varepsilon_m$, $\sum_{i=1}^m g_i = 1$ and some $E > 0$ we have the system of equations

$$\sum_{i=1}^m x_i = 1, \tag{35}$$
$$\sum_{i=1}^m \varepsilon_i x_i = E.$$

where

$$x_i > 0,\ i = 1, \ldots, m, \tag{36}$$

Let $\overline{g\varepsilon} = \frac{1}{m}\sum_{i=1}^m g_i\varepsilon_i$ then we following have lemmas

Lemma 3. *For the system of equations (35) let*

$$x_i = \frac{g_i}{e^{\lambda\varepsilon_i+\nu}},\ i = 1, \ldots, m. \tag{37}$$

where λ and ν are some parameters.

Then

i)	*if $E = \overline{g\varepsilon}$,*	*then the solution is $(\lambda, \nu) = (0, 0)$ and is unique,*
ii)	*if $\varepsilon_1 < E < \overline{g\varepsilon}$,*	*then for $\lambda > 0$ the solution exists and is unique,*
iii)	*if $\overline{g\varepsilon} < E < \varepsilon_m$,*	*then for $\lambda < 0$ the solution exists and is unique,*
iv)	*if $E \notin (\varepsilon_1, \varepsilon_m)$,*	*then the solution does not exists.*

Proof. We start with proof of uniqueness of the solution (λ, ν). First let us assume the solution exists. Then we substitute (37) into (35) and get

$$1 = \sum_{i=1}^m \frac{g_i}{e^{\lambda\varepsilon_i+\nu}}, \tag{38}$$

$$E = \sum_{i=1}^m \frac{g_i\varepsilon_i}{e^{\lambda\varepsilon_i+\nu}}, \tag{39}$$

From the first equation we have

$$\nu = \log\left(\sum_{i=1}^m g_i e^{-\lambda\varepsilon_i}\right). \tag{40}$$

Note that the function $\nu = \nu(\lambda)$ is strictly decreasing, hence one-to-one.
Next we substitute (40) into second equation of (38) and obtain

$$E = \frac{\sum_{i=1}^m g_i \varepsilon_i e^{-\lambda \varepsilon_i}}{\sum_{i=1}^m g_i e^{-\lambda \varepsilon_i}}.$$

Let us show that $E = E(\lambda)$ is ono-to-one function.
We calculate its derivative,

$$E'(\lambda) = \frac{-\sum_{i=1}^m g_i \varepsilon_i^2 e^{-\lambda\varepsilon_i}\left(\sum_{i=1}^m g_i e^{-\lambda\varepsilon_i}\right) + \left(\sum_{i=1}^m \varepsilon_i g_i e^{-\lambda\varepsilon_i}\right)^2}{(\sum_{i=1}^m g_i e^{-\lambda\varepsilon_i})^2},$$

and represent as the difference of the expected values,

$$E'(\lambda) = E[\mathcal{E}]^2 - E[\mathcal{E}^2],$$

where $\mathcal{E}$ is a random variable with range $\Omega_{\mathcal{E}} = \{\varepsilon_1, \ldots, \varepsilon_m\}$ and pdf $f(\varepsilon_i) = g_i e^{-\lambda\varepsilon_i} / \sum_{j=1}^m g_i e^{-\lambda\varepsilon_j}$.
By the Cauchy-Schwartz inequality $E'(\lambda)$ is strictly negative, therefore $E = E(\lambda)$ is strictly decreasing function, hence one-to-one.
Assuming the solution exists, since $E = E(\lambda)$ and $\nu = \nu(\lambda)$ are one-to-one, the solution (λ, ν) is unique.

Now lets prove the existence of the solution (λ, ν).
The function $E(\lambda)$ as $\lambda \to \pm\infty$ and $\lambda = 0$ takes values

$$\begin{aligned}
E(\lambda) =& \frac{\sum_{i=1}^m \varepsilon_i g_i e^{-\lambda\varepsilon_i}}{\sum_{i=1}^m g_i e^{-\lambda\varepsilon_i}} = \frac{g_1\varepsilon_1 + \sum_{i=2}^m \varepsilon_i g_i e^{-\lambda(\varepsilon_i - \varepsilon_1)}}{g_1 + \sum_{i=2}^m g_i e^{-\lambda(\varepsilon_i - \varepsilon_1)}} \to \varepsilon_1, \\
&\text{as } \lambda \to \infty, \\
E(\lambda) =& \frac{1}{m}\sum_{i=1}^m g_i\varepsilon_i, \quad \text{for } \lambda = 0, \\
E(\lambda) =& \frac{g_m\varepsilon_m + \sum_{i=1}^{m-1} g_i\varepsilon_i e^{-\lambda(\varepsilon_i - \varepsilon_m)}}{g_m + \sum_{i=1}^{m-1} e^{-\lambda(\varepsilon_i - \varepsilon_m)}} \to \varepsilon_m, \quad \text{as } \lambda \to -\infty.
\end{aligned}$$

Since $E = E(\lambda)$ is strictly decreasing, points ε_1 and ε_m are boundaries of its range, hence λ exists only if $E \in (\varepsilon_1, \varepsilon_m)$ and further

$$\begin{aligned}
&\text{if } \varepsilon_1 < E < \overline{g\varepsilon}, && \text{then } \lambda > 0, \\
&\text{if } E = \overline{g\varepsilon}, && \text{then } \lambda = 0, \\
&\text{if } \overline{g\varepsilon} < E < \varepsilon_m, && \text{then } \lambda < 0,
\end{aligned}$$

Now, from (40) we have

$$\begin{aligned}
&\nu(\lambda) \to \infty && \text{as } \lambda \to \infty, \\
&\nu(\lambda) = 0 && \text{for } \lambda = 0, \\
&\nu(\lambda) \to -\infty && \text{as } \lambda \to -\infty.
\end{aligned}$$

hence for any λ parameter ν exists.
Putting the results together we get the lemma. $\square$

Lemma 4. *For the system of equations (35) let*

$$x_i = \frac{g_i}{\lambda\varepsilon_i + \nu}, \; i = 1, \ldots, m. \tag{41}$$

where λ and ν are some parameter.
Then,

i)	*if $E = \overline{g\varepsilon}$,*	*then the solution is $(\lambda, \nu) = (0, 1)$ and is unique,*
ii)	*if $\varepsilon_1 < E < \overline{g\varepsilon}$,*	*then the solution exists and is unique for $\lambda > 0, \lambda < -\nu\varepsilon_1$,*
iii)	*if $\overline{g\varepsilon} < E < \varepsilon_m$,*	*then the solution exists and is unique for $\lambda < 0, \lambda > -\nu\varepsilon_m$,*
iv)	*if $E \neq (\varepsilon_1, \varepsilon_m)$,*	*then the solution does not exists.*

Proof. We start with proof of uniqueness of the solution. First we assume it exists. Then we substitute (41) into (35), and get

$$1 = \sum_{i=1}^{m} \frac{g_i}{\lambda\varepsilon_i + \nu}, \tag{42}$$

$$E = \sum_{i=1}^{m} \frac{\varepsilon_i}{\lambda\varepsilon_i + \nu}. \tag{43}$$

and then we perform a substitution $\nu = \lambda\alpha$ and for $\lambda \neq 0$ we have

$$1 = \sum_{i=1}^{m} \frac{g_i}{\lambda(\varepsilon_i + \alpha)}, \tag{44a}$$

$$E = \sum_{i=1}^{m} \frac{g_i\varepsilon_i}{\lambda(\varepsilon_i + \alpha)}. \tag{44b}$$

From (44a) we obtain

$$\lambda = \sum_{i=1}^{m} \frac{g_i}{\varepsilon_i + \alpha}, \tag{45}$$

Except the singularities at the points $\alpha = -\varepsilon_i, i = 1, \ldots, m$, the function $\lambda = \lambda(\alpha)$ is strictly decreasing.
Note that by (36)

$$\frac{1}{\lambda(\varepsilon_i + \alpha)} > 0, \; i = 1, \ldots, m, \tag{46}$$

hence λ and α can take values

$$\lambda > 0, \; \alpha > -\varepsilon_1, \tag{47}$$

or

$$\lambda < 0, \; \alpha < -\varepsilon_m, \tag{48}$$

and if we define $\lambda = \lambda(\alpha)$ separately for the domains (47) and (48) it is also one-to-one. Next we substitute substituting (45) into (44b) and get

$$E = \frac{\sum_{i=1}^{m} \frac{g_i\varepsilon_i}{\varepsilon_i+\alpha}}{\sum_{i=1}^{m} \frac{g_i}{\varepsilon_i+\alpha}}.$$

Let us show that function $E = E(\alpha)$ is one-to-one.
We calculate its derivative,

$$E'(\alpha) = \frac{-\sum_{i=1}^{m} \frac{g_i\varepsilon_i}{(\varepsilon_i+\alpha)^2}\left(\sum_{i=1}^{m} \frac{g_i}{\varepsilon_i+\alpha}\right) + \sum_{i=1}^{m} \frac{g_i\varepsilon_i}{\varepsilon_i+\alpha}\left(\sum_{i=1}^{m} \frac{g_i}{(\varepsilon_i+\alpha)^2}\right)}{\left(\sum_{i=1}^{m} \frac{g_i}{\varepsilon_i+\alpha}\right)^2}$$

and represent it in terms of the expectations

$$E'(\alpha) = E[\mathcal{E}]E\left[\frac{1}{\mathcal{E}+\alpha}\right] - E\left[\frac{\mathcal{E}}{\mathcal{E}+\alpha}\right], \tag{49}$$

where $\mathcal{E}$ and $(\mathcal{E}+\alpha)^{-1}$ are random variables with ranges

$$\Omega_{\mathcal{E}} = \{\varepsilon_1, \dots, \varepsilon_m\},$$
$$\Omega_{(\mathcal{E}+\alpha)^{-1}} = \left\{\frac{1}{\varepsilon_1+\alpha}, \dots, \frac{1}{\varepsilon_m+\alpha}\right\},$$

both with pdf $f_i = \frac{\frac{g_i}{(\varepsilon_i+\alpha)}}{\sum_{j=1}^{m} \frac{g_i}{\varepsilon_j+\alpha}}$.
Now, setting $g(\mathcal{E}) = \mathcal{E}, h(\mathcal{E}) = \frac{1}{\mathcal{E}+\alpha}$ and use special case of FKG inequality, see Appendix on Probability A.3 for the details,

$$E\left[\frac{\mathcal{E}}{\mathcal{E}+\alpha}\right] < E[\mathcal{E}]E\left[\frac{1}{\mathcal{E}+\alpha}\right],$$

which implies $E'(\alpha)$ is strictly positive, therefore $E = E(\alpha)$ is strictly increasing function. If we define $E(\alpha)$ for the domains (47) and (48) separately, it is also one-to-one. Therefore the parameters λ and α are unique.

Next we prove the existence of λ and α.
Let us start with showing the existence of α. The function $E(\alpha)$ as $\alpha \to -\varepsilon_1$, $\alpha \to -\varepsilon_m$ and as $\alpha \to \pm\infty$ takes values

$$E(\alpha) = \frac{\sum_{i=1}^{m} \frac{g_i\varepsilon_i}{\varepsilon_i+\alpha}}{\sum_{i=1}^{m} \frac{g_i}{\varepsilon_i+\alpha}} = \frac{\frac{1}{\alpha}\sum_{i=1}^{m} \frac{g_i\varepsilon_i}{\frac{\varepsilon_i}{\alpha}+1}}{\frac{1}{\alpha}\sum_{i=1}^{m} \frac{g_i}{\frac{\varepsilon_i}{\alpha}+1}} = \frac{1}{m}\sum_{i=1}^{m} g_i\varepsilon_i, \qquad \text{as} \quad \alpha \to \pm\infty, \tag{50}$$

$$E(\alpha) = \frac{\sum_{i=1}^{m} \frac{g_i\varepsilon_i(\varepsilon_1+\alpha)}{\varepsilon_i+\alpha}}{\sum_{i=1}^{m} \frac{g_i(\varepsilon_{im}+\alpha)}{\varepsilon_i+\alpha}} = \frac{g_1\varepsilon_1 + \sum_{i=2}^{m} \frac{g_i\varepsilon_i(\varepsilon_1+\alpha)}{\varepsilon_i+\alpha}}{g_1 + \sum_{i=2}^{m} \frac{g_i(\varepsilon_1+\alpha)}{\varepsilon_i+\alpha}} = \varepsilon_1, \qquad \text{as} \quad \alpha \to -\varepsilon_1, \tag{51}$$

$$E(\alpha) = \frac{\sum_{i=1}^{m} \frac{g_i\varepsilon_i(\varepsilon_m+\alpha)}{\varepsilon_i+\alpha}}{\sum_{i=1}^{m} \frac{g_i(\varepsilon_m+\alpha)}{\varepsilon_i+\alpha}} = \frac{g_m\varepsilon_m + \sum_{i=1}^{m-1} \frac{g_i\varepsilon_i(\varepsilon_m+\alpha)}{\varepsilon_i+\alpha}}{g_m + \sum_{i=m-1}^{m} \frac{g_i(\varepsilon_m+\alpha)}{\varepsilon_i+\alpha}} = \varepsilon_m, \qquad \text{as} \quad \alpha \to -\varepsilon_m. \tag{52}$$

The equation (50) implies that $E(\alpha) \neq \frac{1}{m}\sum_{i=1}^{m} g_i\varepsilon_i$, however if we take original system of equations, i.e. (42), then for $(\lambda, \nu) = (0, 1)$ we have $E = \frac{1}{m}\sum_{i=1}^{m} g_i\varepsilon_i$. Taking that into account and equations (51), (52) we get that α exists if $E \in (\varepsilon_1, \varepsilon_m)$.
Further, since function $E = E(\alpha)$ is strictly decreasing,

if $\varepsilon_1 < E < \overline{g\varepsilon}$,	then $\alpha > -\varepsilon_1$,
if $E = \overline{g\varepsilon}$,	then $\alpha \to \pm\infty$,
if $\overline{g\varepsilon} < E < \varepsilon_m$,	then $\alpha < -\varepsilon_m$.

Now from (45)

$$
\begin{array}{ll}
\lambda(\alpha) \to \infty & \text{as } \alpha \to -\varepsilon_1 \\
\lambda(\alpha) \to 0 & \text{as } \alpha \to \pm\infty, \\
\lambda(\alpha) \to -\infty & \text{as } \alpha \to -\varepsilon_m,
\end{array}
$$

hence if α exists the parameter λ also exists.
Since $\nu = \frac{\alpha}{\lambda}$, λ and α exists and are unique, then the parameter ν also exists and is unique.
Putting the results together we get the outcome of the lemma. □

Lemma 5. *For the system of equations (35) let*

$$x_i = \frac{g_i}{e^{\lambda\varepsilon_i+\nu} - 1}, \; i = 1, \ldots, m. \tag{53}$$

where λ and ν are some parameters.
Then

$$
\begin{array}{lll}
i) & \text{if } E = \overline{g\varepsilon}, & \text{then the solution is } (\lambda,\nu) = (0, \log 2), \\
ii) & \text{if } \varepsilon_1 < E < \overline{g\varepsilon}, & \text{then } \lambda > 0, \lambda < -\nu\varepsilon_1, \\
iii) & \text{if } \overline{g\varepsilon} < E < \varepsilon m, & \text{then } \lambda < 0, \lambda > -\nu\varepsilon_m, \\
iv) & \text{if } E \notin (\varepsilon_1, \varepsilon_m), & \text{then the solution does not exists.}
\end{array}
$$

Proof. As in this case the parameters λ and ν in (53) cannot be factorized w provide the proof without full rigor regarding existence and uniqueness of parameters. We start with proof of uniqueness of the solution. First we assume it exists. Then we substitute (53) into (35), and get

$$1 = \sum_{i=1}^{m} \frac{g_i}{e^{\lambda\varepsilon_i+\nu} - 1}, \tag{54}$$

$$E = \sum_{i=1}^{m} \frac{\varepsilon_i}{e^{\lambda\varepsilon_i+\nu} - 1}. \tag{55}$$

and then we perform a substitution $\nu = \lambda\alpha$ and for $\lambda \neq 0$ we have

$$1 = \sum_{i=1}^{m} \frac{g_i}{e^{\lambda(\varepsilon_i+\alpha)} - 1},$$

$$E = \sum_{i=1}^{m} \frac{g_i\varepsilon_i}{e^{\lambda(\varepsilon_i+\alpha)} - 1},$$

Note that by (36)

$$\frac{g_i}{e^{\lambda(\varepsilon_i+\alpha)} - 1} > 0, \; i = 1, \ldots, m,$$

hence λ and α can take values

$$\lambda > 0, \; \alpha > -\varepsilon_1, \tag{57}$$

or

$$\lambda < 0, \; \alpha < -\varepsilon_m, \tag{58}$$

For $\lambda = 0$ from the system of equation (54) we get that $\nu = \log 2$ and $E = \overline{g\varepsilon}$. Then we get that the solution exists if $E \in (g_{im}\varepsilon_{im}, g_{iM}\varepsilon_{iM})$ as from the second equation of (54) we have

$$\varepsilon_1 = \varepsilon_1 \sum_{i=1}^{m} \frac{g_i}{e^{\lambda\varepsilon_i+\nu}-1} < \sum_{i=1}^{m} \frac{g_i\varepsilon_i}{e^{\lambda(\varepsilon_i+\alpha)}-1},$$
$$\varepsilon_m = \varepsilon_m \sum_{i=1}^{m} \frac{g_i}{e^{\lambda\varepsilon_i+\nu}-1} > \sum_{i=1}^{m} \frac{g_i\varepsilon_i}{e^{\lambda(\varepsilon_i+\alpha)}-1},$$

i.e. the weighted sum cannot exceed its highest element or be smaller than lowest.

(???) Since $E = E(\lambda, \nu)$ is a strictly decreasing function w.r.t variable λ we have that for $\varepsilon_1 < E < \overline{g\varepsilon}$ the corresponding parameters λ and ν are in the regime given by (57). For the values of E in $\varepsilon_m > E > \overline{g\varepsilon}$ we have the other regime (58). Putting together the outcomes we get the final result. □

A Analysis

Theorem 3 (Taylor's expansion)**.** *Suppose that f is a real function on the nontrivial convex closed set $A \in \mathbb{R}^m$, n is a positive integer, $f(n-1)$ is continuous on A, $f^{(n)}(t)$ exists for every $t \in A$. Then there exists a point x_θ between x^* and x, such that*

$$f(x) = f(x^*) + \sum_{k=1}^{n-1} \frac{1}{k!} \sum_{i_1,i_2,\dots,i_k=1}^{m} \frac{\partial^k f(x^*)}{\partial x_{i_1}\partial x_{i_2}\dots\partial x_{i_k}}(x_{i_1}-x^*_{i_1})(x_{i_2}-x^*_{i_2})\dots(x_{i_k}-x^*_{i_k})+$$
$$+\frac{1}{n!}\sum_{i_1,i_2,\dots,i_n=1}^{m} \frac{\partial^m f(x_\theta)}{\partial x_{i_1}\partial x_{i_2}\dots\partial x_{i_m}}(x_{i_1}-x^*_{i_1})(x_{i_2}-x^*_{i_2})\dots(x_{i_m}-x^*_{i_m}), \tag{59}$$

where x_θ can be represented, $x_\theta = x^* + \theta(x - x^*)$, $0 < \theta < 1$.

The $k-th$ in the Taylor's theorem can be represented

$$\sum_{i_1,i_2,\dots,i_k=1}^{m} \frac{f^k(x)}{\partial x_{i_1}\partial x_{i_2}\dots\partial x_{i_k}}(x_{i_1}-x^*_{i_1})\dots(x_{i_k}-x^*_{i_k}) = D^k f(x)(x-x^*)^{\otimes k}, \tag{60}$$

where $\otimes$ is tensor product, $D : f \to \nabla f$ differentiation operator and $D\otimes D\otimes\dots\otimes D = D^{\otimes k}$ and $D^{\otimes k} f(x)|_{x=x_\theta} = D^{\otimes k} f(x_\theta)$ is an operater such that $D^{\otimes k} f(x_\theta) : \mathbb{R}^{m^k} \to \mathbb{R}$.
Hence, for $n \geq 4$ expansion (59) can also be represented as

$$\begin{aligned} f(x) =& f(x^*) + Df(x^*)^T(x-x^*) + (x-x^*)^T D^2 f(x^*)(x-x^*)+ \\ &+\sum_{k=3}^{n-1}\frac{1}{k!}D^{\otimes k} f(x^*)(x-x^*)^{\otimes k} + \frac{1}{n!}D^{\otimes n} f(x_\theta)(x-x^*)^{\otimes n}, \end{aligned}$$

where $0 < \theta < 1$.
Since operator $D^{\otimes n} f(x_\theta)$ is finite dimensional, hence it is bounded and we have

$$\left|D^{\otimes n} f(x_\theta)(x-x^*)^{\otimes n}\right| \leq \|D^{\otimes n} f(x_\theta)\||(x-x^*)^{\otimes n}|. \tag{61}$$

Were we define
$$\|D^{\otimes n}f(x_\theta)\| = \sup_{x\in A}\frac{|D^{\otimes k}f(x_\theta)(x-x^*)^{\otimes k}|}{|(x-x^*)^{\otimes k}|}.$$
which is a norm of the operator. Further, by the definition of the tensor product we have that
$$< x\otimes x, x\otimes x >=< x, x >< x, x >$$
which implies $|x^{\otimes k}| = |x|^k$. Putting that fact together with (61) we get that n-th term in Taylor's Theorem is bounded by
$$\left|D^{\otimes n}f(x_\theta)(x-x^*)^{\otimes n}\right| \le \|D^{\otimes n}f(x_\theta)\||x-x^*|^n. \tag{62}$$
If the operator $D^{\otimes k}f(x)$ is invertible, then we have $x = D^{\otimes k}f(x)^{-1}y$ and define
$$\|D^2f(x)^{-1}\| = \sup_{y\in B}\frac{|D^{\otimes k}f(x)^{-1}y|}{|y|}$$
where B is a range of the considered operator. Then we have inequality
$$\left|D^{\otimes k}f(x)^{-1}y\right| \le \|D^2f(x)^{-1}\||y|,$$
which we transform taking into account relation $x = D^{\otimes k}f(x)^{-1}y$
$$\frac{1}{\|D^2f(x)^{-1}\|}|x| \le \left|D^{\otimes k}f(x_\theta)x\right|,$$
setting $x = (x-x^*)^{\otimes k}$ and including $|(x-x^*)^{\otimes k}| = |x-x^*|^k$ we get
$$\frac{|x-x^*|^k}{\|D^2f(x)^{-1}\|} \le \left|D^{\otimes k}f(x_\theta)(x-x^*)^{\otimes k}\right|, \tag{63}$$
which is a lower bound for n-th term in the Taylor expansion.
Analogically to Theorem 1 we can define follwing Taylor's expansion
$$\begin{aligned}\frac{\partial f(x)}{\partial x_j} &= \frac{\partial f(x^*)}{\partial x_j} + \sum_{k=1}^{n-1}\frac{1}{k!}\sum_{i_1,i_2,\ldots,i_k=1}^{m}\frac{\partial^{k+1}f(x^*)}{\partial x_j\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_k}}(x_{i_1}-x^*_{i_1})(x_{i_2}-x^*_{i_2})\ldots(x_{i_k}-x^*_{i_k})+\\ &+\frac{1}{n!}\sum_{i_1,i_2,\ldots,i_n=1}^{m}\frac{\partial^{m+1}f(x_\theta)}{\partial x_j\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_m}}(x_{i_1}-x^*_{i_1})(x_{i_2}-x^*_{i_2})\ldots(x_{i_m}-x^*_{i_m}),\end{aligned} \tag{64}$$
where x_θ is somewhere between x and x^*. The vector composed of elements of above expansion for $j = 1,\ldots,m$ can be defined as
$$Df(x) = Df(x^*) + D^2f(x^*)(x-x^*) + \sum_{k=2}^{n-1}\frac{1}{k!}D^{\otimes(k+1)}f(x^*)(x-x^*)^{\otimes k} + \frac{1}{n!}D^{\otimes(n+1)}f(x_\theta)(x-x^*)^{\otimes n},$$
where we define operator $D^{\otimes(k+1)}f(x) : \mathbb{R}^{m^k} \to \mathbb{R}^m$ for $k \ge 1$ as
$$\left[D^{\otimes(k+1)}f(x)(x-x^*)^{\otimes k}\right]_j = \sum_{i_1,i_2,\ldots,i_k=1}^{m}\frac{\partial^{k+1}f(x)}{\partial x_j\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_k}}(x_{i_1}-x^*_{i_1})(x_{i_2}-x^*_{i_2})\ldots(x_{i_k}-x^*_{i_k}),$$

for $j = 1, \ldots, m$.
Analogically to the bounds of the remainder of the previous Taylor's expansion (62) and (63) we have for $D^{\otimes(k+1)} f(x)$

$$\left| D^{\otimes(k+1)} f(x_\theta)(x - x^*)^{\otimes k} \right| \leq \|D^{\otimes(k+1)} f(x_\theta)\| |x - x^*|^k. \tag{65}$$

Moreover, if $D^{\otimes(k+1)} f(x)$ is invertible we have

$$\frac{|x - x^*|^k}{\|D^{\otimes(k+1)} f(x_\theta)^{-1}\|} \leq \left| D^{\otimes(k+1)} f(x_\theta)(x - x^*)^{\otimes k} \right|.$$

Further, we define another Taylor's expansion, analogical to one in Theorem 1 as

$$\begin{aligned} \frac{\partial^2 f(x)}{\partial x_j \partial x_l} = \frac{\partial^2 f(x^*)}{\partial x_j \partial x_l} + \sum_{k=1}^{n-1} \frac{1}{k!} \sum_{i_1, i_2, \ldots, i_k = 1}^{m} \frac{\partial^{k+2} f(x^*)}{\partial x_j \partial x_l \partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_k}} (x_{i_1} - x^*_{i_1})(x_{i_2} - x^*_{i_2}) \ldots (x_{i_k} - x^*_{i_k}) + \\ + \frac{1}{n!} \sum_{i_1, i_2, \ldots, i_n = 1}^{m} \frac{\partial^{m+2} f(x_\theta)}{\partial x_j \partial x_l \partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_m}} (x_{i_1} - x^*_{i_1})(x_{i_2} - x^*_{i_2}) \ldots (x_{i_m} - x^*_{i_m}), \end{aligned} \tag{66}$$

where again, x_θ is somewhere between x and x^*. The matrix composed of elements of above expansion for $j = 1, \ldots, m,\ l = 1, \ldots, m$ can be defined as

$$D^2 f(x) = D^2 f(x^*) + \sum_{k=1}^{n-1} \frac{1}{k!} D^{\otimes(k+2)} f(x^*)(x - x^*)^{\otimes k} + \frac{1}{n!} D^{\otimes(n+2)} f(x_\theta)(x - x^*)^{\otimes n},$$

where we define $D^{\otimes(k+2)} f(x) : \mathbb{R}^{m^k} \to \mathbb{R}^{m^2}$ for $k \geq 1$ as

$$\left[D^{\otimes(k+2)} f(x)(x - x^*)^{\otimes k} \right]_{j,l} = \sum_{i_1, i_2, \ldots, i_k = 1}^{m} \frac{\partial^{k+2} f(x)}{\partial x_j \partial x_l \partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_k}} (x_{i_1} - x^*_{i_1})(x_{i_2} - x^*_{i_2}) \ldots (x_{i_k} - x^*_{i_k}),$$

for $j = 1, \ldots, m,\ l = 1, \ldots, m$.
Then again we have that

$$\left| D^{\otimes(n+2)} f(x_\theta)(x - x^*)^{\otimes n} \right| \leq \|D^{\otimes(n+2)} f(x_\theta)(x - x^*)^{\otimes n}\| |x - x^*|^n,$$

and if $D^{\otimes(n+2)} f(x_\theta)(x - x^*)^{\otimes n}$ is invertible we have

$$\frac{|x - x^*|^n}{\|D^{\otimes(n+2)} f(x_\theta)(x - x^*)^{\otimes n}\|} \leq \left| D^{\otimes(n+2)} f(x_\theta)(x - x^*)^{\otimes n} \right|,$$

where $n \geq 1$.

B Asymptotic theory and approximation

Let $f : A \to \mathbb{R}$ be a continuous function and $A = (a, \infty)$ for some a.

Definition 1 (Big O)**.** *The function f is of order O of the function $g : A \to \mathbb{R}$ as $x \to \infty$ if there exists is a constant $K > 0$ and $x_K \in A$ such that for all $x > x_K$*

$$|f(x)| \leq K|g(x)|,$$

and we write it symbolically

$$f(x) = O(g(x)),\ x \to \infty.$$

Definition 2 (Small o)**.** *The function f is of order o of the function $g : A \to \mathbb{R}$ as $x \to \infty$ if for all $K > 0$ there exists $x_K \in A$ such that for all $x > x_K$*

$$|f(x)| \leq K|g(x)|,$$

and we write it symbolically

$$f(x) = o(g(x)),\ x \to \infty.$$

Definition 3 (Asymptotic equivalence)**.** *The functions f and $g : A \to \mathbb{R}$ are asymptotically equivalent as $x \to \infty$ if for all $K > 0$ there exists $x_K \in A$ such that for all $x > x_K$, $f(x) \neq 0$ and $g(x) \neq 0$ and*

$$\left|\frac{f(x)}{g(x)} - 1\right| \leq K,$$

and we write it symbolically

$$f(x) \sim g(x),\ x \to \infty.$$

Definition 4 (Asymptotic expansion)**.** *The formal power series $\sum_{n=0}^{\infty} a_n x^{-n}$ is an asymptotic power series expansion of f, as $x \to \infty$ if for all $m \in \mathbb{N}$*

$$f(x) = \sum_{n=0}^{m} a_n x^{-n} + O(x^{-(m+1)}),\ x \to \infty, \tag{67}$$

and we write it symbolically

$$f(x) \sim \sum_{n=0}^{\infty} a_n x^{-n},\ x \to \infty.$$

If first few coefficients of power series are known then we write

$$f(x) \sim a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \dots,\ x \to \infty.$$

Furthermore (67) can equivalently be represented as

$$f(x) = \sum_{n=0}^{m} a_n x^{-n} + \sigma(x), \tag{68}$$
$$\sigma(x) = O(x^{-(m+1)}), x \to \infty.$$

Let the function $\Gamma(\lambda) = (\lambda - 1)!$ where ! is a usual factorial and $\lambda \in \mathbb{N}$. For $\lambda \in \mathbb{R}$ it is defined through its integral form

$$\Gamma(\lambda) = \int_0^{\infty} t^{\lambda-1} e^{-t} dt.$$

Theorem 4 (Gamma function approximation)**.** *The Gamma function* $\Gamma(\lambda)$ *can be approximated*

$$\Gamma(\lambda) \sim e^{-\lambda}\lambda_{\lambda}\left(\frac{2\pi}{\lambda}\right)^{1/2}\left[1+\frac{1}{12\lambda}+\frac{1}{288\lambda^2}+\dots\right].$$

Proof. See **?**, p.60. $\square$

Lemma 6. *Given inequalities*

$$A_i - B_i = C_i, \;\; i = 1, \dots, m, \tag{69}$$

$$\tag{70}$$

where $m \in \mathbb{N}$ *following inequalities holds*

$$\prod_{i=1}^{m} A_i - \prod_{i=1}^{m} B_i = \prod_{i=1}^{m} C_i + \sum_{j=1}^{m-1} \sum_{C_m^{m-j,j}} \prod_{k=1}^{j} \prod_{l=j+1}^{m} C_{i_k} B_{i_l},$$

Proof. We start by introducing equality

$$\prod_{i=1}^{m} A_i = (A_1 - B_1)\prod_{i=2}^{m} A_i + B_1 \prod_{i=2}^{m} A_i,$$

which we obtained by adding and deducting $B_1 \prod_{i=2}^{m} A_i$ to $\prod_{i=1}^{m} A_i$.
Then again, we add and deduce, but this time $B_1 B_2 \prod_{l=3}^{m} A_l$ and $(A_1 - B_1)B_2 \prod_{l=3}^{m} A_l$ and get

$$\prod_{i=1}^{m} A_i = (A_1-B_1)(A_2-B_2)\prod_{i=3}^{m} A_i + (A_1-B_1)B_2\prod_{i=3}^{m} A_i + B_1(A_2-B_2)\prod_{i=3}^{m} A_i + B_1B_2\prod_{i=3}^{m} A_i.$$

We repeat that step until all A_i's in the products are replaced by $(A_i - B_i)$, which eventually leads to the equation

$$\prod_{i=1}^{m} A_i - \prod_{i=1}^{m} B_i = \prod_{i=1}^{m}(A_i - B_i) + \sum_{j=1}^{m-1} \sum_{C_m^{m-j,j}} \prod_{k=1}^{j} \prod_{l=j+1}^{m} (A_{i_k} - B_{i_k})B_{i_l} \tag{71}$$

where $\sum_{C_m^{m-j,j}}$ is a sum over possible arrangements of the elements of the set $\{1, 2, \dots, m\}$ into two groups, where elements does not repeat and within the group the order does not matter. First group is of the size $m - j$ and second j and their elements correspond, respectively, to the indecies $i_k, k = 1, \dots, m - j$ and $i_l, l = m - j + 1 \dots, m$.
Then we substitute (69) and obtain

$$\prod_{i=1}^{m} A_i - \prod_{i=1}^{m} B_i = \prod_{i=1}^{m} C_i + \sum_{j=1}^{m-1} \sum_{C_m^{m-j,j}} \prod_{k=1}^{j} \prod_{l=j+1}^{m} C_{i_k} B_{i_l},$$

which is our result. $\square$

Lemma 7. *Given equalities*

$$A_1 - B_1 = \sigma_1, \tag{72}$$

$$A_2 - B_2 = \sigma_2, \tag{73}$$

where $\sigma_1, \sigma_2 \in \mathbb{R}$ *are some constants, following equality holds*

$$\frac{A_1}{A_2} - \frac{B_1}{B_2} = -\frac{B_1\sigma_2}{B_2(B_2\sigma_2)} + \frac{\sigma_1}{B_2 + \sigma_2}.$$

Proof. First we divide (72) by (73) and get

$$\frac{A_1}{A_2} = \frac{B_1 + \sigma_1}{B_2 - \sigma_2}, \tag{74}$$

and then we transform LHS

$$\frac{B_1 + \sigma_1}{B_2 + \sigma_2} = \frac{B_1}{B_2 + \sigma_2} + \frac{\sigma_1}{B_2 + \sigma_2},$$

then add and deduce B_1/B_2 and perform some manipulations

$$\begin{aligned}\frac{B_1}{B_2 + \sigma_2} + \frac{\sigma_1}{B_2 + \sigma_2} &= \frac{B_1}{B_2} + \frac{B_1}{B_2 + \sigma_2} - \frac{B_1}{B_2} + \frac{\sigma_1}{B_2 + \sigma_2} = \\ &= \frac{B_1}{B_2} + \frac{B_1 B_2}{B_2(B_2 + \sigma_2)} - \frac{B_1(B_2 + \sigma_2)}{B_2(B_2 + \sigma_2)} + \frac{\sigma_1}{B_2 + \sigma_2} = \\ &= \frac{B_1}{B_2} - \frac{B_1 \sigma_2}{B_2(B_2 + \sigma_2)} + \frac{\sigma_1}{B_2 + \sigma_2},\end{aligned}$$

and then we put it back to (74), change the side of B_1/B_2 and obtain final result

$$\frac{A_1}{A_2} - \frac{B_1}{B_2} = -\frac{B_1 \sigma_2}{B_2(B_2 + \sigma_2)} + \frac{\sigma_1}{B_2 + \sigma_2}.$$

□

C Theory of Optimization

Definition 5 (Optimization problem in standard form). *An optimization problem in standard form has the form*

$$\begin{aligned} &\textit{minimize } f_0(x), \\ &\textit{subject to } f_i(x) \leq 0, i = 1, \ldots, m, \\ &\qquad\qquad h_i(x) = 0, i = 1, \ldots, p, \end{aligned}$$

The vector $x \in \mathbb{R}^m$ is the optimization variable of the problem. The function $f_0(x)$ the objective function. The inequalities $f_i(x) \leq 0$ are called inequality constraints and equalities $h_i(x) = 0$ are equality constraints.
The domain for which on which objective function and all constraints is defined $\mathcal{D}$ we call a domain of the optimization problem. Any point $x \in \mathcal{D}$ is feasible if it satisfies the all the constraints.
Furthermore, the optimal value p^* is defined as

$$p^* = \inf\{f_0(x) | f_i(x) \leq 0, i = 1, \ldots, m,\ h_i(x) = 0, i = 1, \ldots, p\}.$$

and x^* is an optimal point (vector) , if x^* is feasible and $f_0(x^*) = p^*$.

Definition 6 (Convex problem). *An optimization problem is convex problem if it is of the form*

$$\begin{aligned} &\textit{minimize } f_0(x). \\ &\textit{subject to } f_i(x) \leq 0, i = 1, \ldots, m, \\ &\qquad\qquad h_i(x) = 0, i = 1, \ldots, p, \end{aligned}$$

where $f_0, f_1, \ldots, f_m$ are convex. Furthermore requirements must be met

- *optimized function f_0 is convex,*
- *inequality constraint functions must be convex,*
- *equality constraints must be affine.*

Definition 7 (Affine hull)**.** *We define affine hull by set of all affine combinations of points in some set $A \subseteq \mathbb{R}^m$ is called the affine hull of A and denoted by aff(A):*

$$\mathit{aff}(A) = \{\theta_1 x_1 + \ldots + \theta_k x_k \mid x_1, \ldots, x_k \in A\}$$

Definition 8 (Relative interior)**.** *We define relative interior of the set C as*

$$\mathit{relint}(C) = \{x \in C \mid \exists_{r>0} B(x, r) \cap \mathit{aff}A \subseteq C\}$$

Theorem 5 (weak Slater's condition)**.** *The Slater's condition hold if optimization problem is convex and there exists $x \in \mathit{relint}(\mathcal{D})$ with*

$$f_i(x) \leq 0, \ i = 1, \ldots, k, \ f_i(x) < 0, \ i = k+1, \ldots, m.$$

where f_i are inequality constraints and first k of them are affine and relint($\mathcal{D}$) is relative interior of the domain. Moreover if Slater's conditions hold then optimal vector (λ^*, ν^*) exists and strong duality occurs.

Proof. See Boyd and Vandenberghe [2004], p.227. □

Theorem 6 (KKT conditions for convex problem)**.** *The Karush-Kuhn-Tucker (KKT) conditions are*

$$\begin{aligned} f_i(x^*) &\leq 0, \\ h_i(x^*) &= 0, \\ \tilde{\lambda}_i &\geq 0, \\ \tilde{\lambda}_i f_i(x^*) &= 0, \\ \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{m} \nu_i^* \nabla h_i(x^*) &= o. \end{aligned}$$

For any convex problem with differentiable objective and constraint functions, any points that satisfy KKT conditions are primal and dual optimal and strong duality holds.

Furthermore, if Slater's conditions holds then KKT are necessary and sufficient conditions for the optimality.

Proof. See Boyd and Vandenberghe [2004], p.244. □

References

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

Vassili N. Kolokoltsov and Tomasz M. Lapinski. Weak law of large numbers and central limit theorem for probabilistic system with given entropy. *http://arxiv.org/abs/1501.06376*, 2015.

Victor P. Maslov. Nonlinear averaging axioms in financial mathematics and stock prices. *Theory Probability Applications*, 48(4):723–733, 2004.

Victor P. Maslov. On a general theorem of set theory leading to the gibbs, bose-einstein, and pareto distributions as well as to the zipf-mandelbrot law for the stock market. *Mathematical Notes*, 78(6):807–813, 2005a.

Victor P. Maslov. Nonlinear averages in economics. *Mathematical Notes*, 78(3):347–363, 2005b.